

On the Relationships among Optimal Symmetric Fix-Free Codes

S. M. Hossein Tabatabaei Yazdi
Texas A&M University
smhtyazdi@neo.tamu.edu

Serap A. Savari
Texas A&M University
savari@ece.tamu.edu

Abstract

Symmetric fix-free codes are prefix condition codes in which each codeword is required to be a palindrome. Their study is motivated by the topic of joint source-channel coding. Although they have been considered by a few communities they are not well understood. In earlier work we used a collection of instances of Boolean satisfiability problems as a tool in the generation of all optimal binary symmetric fix-free codes with n codewords and observed that the number of different optimal codelength sequences grows slowly compared with the corresponding number for prefix condition codes. We demonstrate that all optimal symmetric fix-free codes can alternatively be obtained by sequences of codes generated by simple manipulations starting from one particular code. We also discuss simplifications in the process of searching for this set of codes.

1. Introduction

Shannon's pioneering work on information theory [15] establishes that source and channel encoding can be separated without a loss of performance assuming infinite blocklengths are permitted. However, that result does not apply to real transmission situations with complexity and latency constraints, and there is therefore an interest in joint source-channel coding and decoding techniques. Many video, audio, and image standards use prefix condition codes. It is therefore interesting to devise prefix condition codes with additional constraints which result in binary encodings of data with increased immunity to noise prior to channel encoding. For example, *fix-free* or reversible variable length codes (see, e.g., [14], [7], [4], [16]) are prefix condition codes in which no codeword is the suffix of another codeword, and they are components of the video standards H.264 and MPEG-4 [17], [9], [20], [10].

Our focus in this paper is upon a subclass of fix-free codes known as *symmetric* fix-free codes [16]. Here each codeword must be a palindrome. Symmetric fix-free codes were found [2] to be preferable to other fix-free codes for joint source-channel coding. They are also easier to study because a collection of palindromes which satisfies the prefix condition automatically satisfies the suffix condition [16], [18], [12]. Nevertheless, although they have also been studied in [3], [17], [19], [1], [8], [13] they are not well-understood. For example, there is no exact counterpart to the Kraft inequality/equality for symmetric fix-free codes, although [16], [18], [12], [13] discuss some simple nonexhaustive necessary and sufficient conditions for the codeword lengths of such codes. In [12], [1], [13] we convert the problem of determining the existence of a symmetric fix-free code with given codeword lengths into a Boolean satisfiability problem and offer

branch-and-bound algorithms to find the set of optimal codes for all memoryless sources, i.e., codes which minimize the average codeword length among all symmetric fix-free codes for some choice of source probabilities. For a given source its optimal code can be found by calculating the expected codeword length for each of the optimal codelength sequences and choosing the corresponding optimal code. In [1], [13] we show that the number of sorted and nondecreasing optimal codelength sequences for binary symmetric fix-free codes with n codewords appears to grow very slowly with n compared with the corresponding exponential growth [6] for binary prefix condition codes (see the appendix). Therefore, when n is not too large it appears to be feasible to calculate and store all optimal codes and to choose the best among them for a given application. The paper [8] proposes an A^* -based algorithm for a different way to obtain an optimal symmetric fix-free code for a given source, but this procedure does not offer much mathematical insight about optimal codes. The existing understanding about optimal codes is very limited.

Although solving instances of Boolean satisfiability problems can be one component in the generation of optimal codes, we propose in Section 3 a completely different derivation of them. Our inspiration comes from a paper [11] which shows that the space of all sorted and non-decreasing sequences of codeword lengths of optimal binary prefix condition codes forms a lattice called the *imbalance* lattice. Among the length sequences which satisfy the Kraft inequality with equality, $(1, 2, 3, \dots, n-1, n-1)$ is considered to be the most imbalanced because it corresponds to the largest sum of codeword lengths. The authors of [11] describe a basic operation on three values of a codeword length sequence which when repeated enough times will transform the most imbalanced codeword length sequence into an arbitrary sorted and non-decreasing optimal codeword length sequence.

We will not work here with length sequences but instead with the binary codes themselves. Although the optimal codes do not form a lattice we will see that they can each be attained from the repetition of a basic operation which eventually transforms the most “imbalanced” optimal code into an arbitrary optimal code. (The basic operation here is completely different from that of [11], and the number of codewords it will affect in one application depends on several factors.) The following results from [13] show that the most imbalanced optimal symmetric fix-free code is $\{0, 11, 101, 1001, \dots\}$ with length sequence $(1, 2, \dots, n)$.

Proposition 1: [13, Prop. 2.2] The code $\{0, 11, 101, 1001, \dots\}$ with $n \geq 3$ codewords is in the set of optimal symmetric fix-free codes with n codewords.

Theorem 2: [13, Thm. 2.5] The sorted and non-decreasing length sequence (l_1, l_2, \dots, l_n) of an optimal binary symmetric fix-free code with n codewords satisfies $l_i \leq n$ for $i \in \{1, 2, \dots, n\}$ and $\sum_{i=1}^n l_i \leq \sum_{i=1}^n i = n(n+1)/2$.

Our initial procedure to generate any optimal symmetric fix-free code will also generate some suboptimal codes. Part of the contribution of Section 4 is to provide simple tests to reduce the number of candidates for optimal codes, and one of these tests can be viewed as a generalization of Theorem 2.

2. Preliminaries

Given a palindrome σ , we define the set of its *neighboring palindromes* $\mathcal{N}(\sigma)$ by $\mathcal{N}(\sigma) = \{\text{palindromes } w: \sigma \text{ is the longest palindrome which is a proper prefix of } w\}$.

For example, $\mathcal{N}(0) = \{00, 010, 0110, \dots\}$. For any string w , let $|w|$ denote the length of w . We will be interested in the following (possibly empty) subset of $\mathcal{N}(\sigma)$

$$\mathcal{N}_n(\sigma) = \{w \in \mathcal{N}(\sigma) : |w| \leq n\}.$$

Note that if we remove a palindrome σ from a symmetric fix-free code, then we can add to the remainder of that code any subset of $\mathcal{N}(\sigma)$ to obtain another symmetric fix-free code with possibly more codewords than the original code.

Observe that for any symmetric fix-free code $C_n = \{c_1, c_2, \dots, c_n\}$, we can define a “complementary” symmetric fix-free code by reversing the bits of each codeword. For $n \geq 3$ any symmetric fix-free code will have at most one codeword consisting of a single bit, so we can assume without loss of generality that $1 \notin C_n$. We will ultimately be concerned with the set \mathbb{O}_n of optimal symmetric fix-free codes C_n with n codewords for which $1 \notin C_n$. However, we begin by considering the larger set \mathbb{S}_n of symmetric fix-free codes C_n with n codewords for which $1 \notin C_n$ and $\max_{1 \leq i \leq n} |c_i| \leq n$.

We will call the symmetric fix-free code $\{0, 11, 101, 1001, \dots\}$ with length sequence $(1, 2, \dots, n)$ the *root* code of length n and label it R_n . We have the following result.

Lemma 3: Any codeword of a symmetric fix-free code $C_n \in \mathbb{S}_n$ has a codeword of R_n as a prefix.

Proof: Let s_i , $i \leq n$, denote the codeword of length i in R_n . All codewords in C_n which begin with a 0 have s_1 as the prefix. All other codewords in C_n begin with a 1, and by assumption, $1 \notin C_n$. Observe that any binary string beginning with a 1 and having length between 2 and n will either have s_i as a prefix for some $2 \leq i \leq n$ or it will be in the set $\{10, 100, 1000, \dots\}$. However, a binary string beginning with a 1 and ending with a 0 is not a palindrome and is therefore not in C_n . \square

3. Relations among Optimal Symmetric Fix-Free Codes

We define two relations \rightarrow and \Rightarrow between codes $S_n, \hat{S}_n \in \mathbb{S}_n$ by

$$S_n \rightarrow \hat{S}_n \quad \text{if there exists } \sigma \in S_n \text{ such that } \hat{S}_n \subseteq S_n \cup \mathcal{N}_n(\sigma) \setminus \{\sigma\}.$$

For this σ we write $S_n \xrightarrow{\sigma} \hat{S}_n$.

$$S_n \Rightarrow \hat{S}_n \quad \text{if there exists } \sigma \in S_n \text{ such that } \hat{S}_n \text{ consists of the shortest } n \text{ words of } S_n \cup \mathcal{N}_n(\sigma) \setminus \{\sigma\}. \text{ For this } \sigma \text{ we write } S_n \xRightarrow{\sigma} \hat{S}_n.$$

We have the following result about \mathbb{S}_n .

Theorem 4: For any code $C_n \in \mathbb{S}_n$ with codeword lengths l_1, l_2, \dots, l_n , there exists an integer $m \leq \sum_{i=1}^n (l_i - 1) = O(n^2)$ and a sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)} \in \mathbb{S}_n$ for which $R_n = S_n^{(0)} \rightarrow S_n^{(1)} \rightarrow S_n^{(2)} \rightarrow \dots \rightarrow S_n^{(m)} = C_n$ and with the property that each codeword of C_n has a prefix in $S_n^{(i)}$ for each $i \in \{0, 1, \dots, m-1\}$. Furthermore, there exists a code $B_n \in \mathbb{S}_n$ for which the preceding sequence requires $m = \Omega(n^{1.5})$ codes.

Proof: Consider the following algorithm to generate the codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)}$:

- 1) $S_n^{(0)} = R_n$; $i = 0$.
- 2) If there exists a codeword $w \in C_n$ which has a proper prefix $\sigma \in S_n^{(i)}$:
 - a) Find the subset $C_n(\sigma)$ of $\mathcal{N}_n(\sigma)$ consisting of the strings which are prefixes of codewords of the code C_n . If there are $\#(\sigma)$ words in $C_n(\sigma)$, then there

is a subset $D^{(i)} \subset S_n^{(i)} \setminus \{\sigma\}$ with $\#(\sigma) - 1$ strings such that no element of $D^{(i)}$ is a prefix of a word in C_n .

b) Set $S_n^{(i+1)} = S_n^{(i)} \cup C_n(\sigma) \setminus \{\{\sigma\} \cup D^{(i)}\}$.

3) $i \leftarrow i + 1$. Goto 2.

We argue inductively that this procedure generates an appropriate sequence of codes. For the basis step, we have seen in Lemma 3 that every element of C_n has a prefix in $R_n = S_n^{(0)}$. For the inductive step, assume that every element of C_n has a prefix in $S_n^{(k)}$ for some $k \geq 0$, and assume $w \in C_n$ has a proper prefix σ in $S_n^{(k)}$. Since $\mathcal{N}_n(\sigma)$ contains the palindromes of length at most n for which σ is the longest proper prefix which is a palindrome, w has a prefix (possibly the full string) which is an element of $\mathcal{N}_n(\sigma)$. That prefix will be a member of $S_n^{(k+1)}$, and we repeat this argument for any other codeword of C_n having σ as a prefix. For each codeword of C_n having a different prefix in $S_n^{(k)}$, we assume that the same prefix will be an element of $S_n^{(k+1)}$. Therefore $S_n^{(k+1)}$ has the desired property.

For an upper bound on m , each application of operation \rightarrow will involve a different choice for the string σ , and each one will be a palindrome which is a proper prefix of at least one codeword. The result follows since each codeword of length l_i , $i \in \{1, 2, \dots, n\}$, has $l_i - 1 \leq n - 1$ proper prefixes.

For the last part, our code B_n will consist of n palindromes of length n which begin with and end with 0. For convenience we assume here that n is even. Since there are $2^{0.5n-1}$ such palindromes, we must have $n \geq 8$. We will describe the code in terms of l clusters of codewords. The first cluster is the all-zero string, which has $n - 1$ proper prefixes all of which are palindromes. The second cluster is a single string with left half 0101.... The new proper prefixes which are palindromes are 010, 01010, ..., and there are $(1/2) \cdot (0.5n - 2 - O(1))$ of them. The third cluster consists of the two strings with left half 0110110110... and left half 00100100.... The new proper prefixes of the left halves of these string which are palindromes are 0110, 00100, 0110110, 00100100, ..., and there are $(2/3) \cdot (0.5n - 3 - O(1))$ of them. Cluster j , $j \in \{2, 3, \dots, l\}$, consists of $j - 1$ strings. The left half of string $k \in \{1, \dots, j - 1\}$ of cluster j is a repetition of the length j string beginning with k zeroes and ending with $j - k$ ones. There are $((j - 1)/j) \cdot (0.5n - j - O(1))$ proper prefixes of the left halves of these strings. Since there are n words in the combination of all clusters, we have that $l = \Omega(\sqrt{n})$, and the number of proper prefixes of all n codewords is $\Omega(n^{1.5})$. \square

We can characterize the set of optimal codes as follows.

Theorem 5: For any code $C_n \in \mathbb{O}_n$ there exist an integer $m = O(n^2)$ and a sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)} \in \mathbb{S}_n$ for which $R_n = S_n^{(0)} \Rightarrow S_n^{(1)} \Rightarrow S_n^{(2)} \Rightarrow \dots \Rightarrow S_n^{(m)} = C_n$.

Proof: By Theorem 4, there exist $m = O(n^2)$, a sequence of codes $C_n^{(1)}, C_n^{(2)}, \dots, C_n^{(m)} \in \mathbb{S}_n$, and palindromes $w_i \in C_n^{(i)}$, $0 \leq i \leq m - 1$, such that $R_n = C_n^{(0)} \xrightarrow{w_0} C_n^{(1)} \xrightarrow{w_1} C_n^{(2)} \xrightarrow{w_2} \dots \xrightarrow{w_{m-1}} C_n^{(m)} = C_n$; i.e.,

$$C_n^{(i+1)} \subseteq C_n^{(i)} \cup \mathcal{N}_n(w_i) \setminus \{w_i\}. \quad (1)$$

Let $k \geq 1$ be the smallest integer for which $C_n^{(k-1)} \not\supseteq C_n^{(k)}$, and let $S_n^{(k)}$ denote the choice of the shortest n strings in $C_n^{(k-1)} \cup \mathcal{N}_n(w_{k-1}) \setminus \{w_{k-1}\}$ which has maximum overlap

with $C_n^{(k)}$. Therefore, for any $c \in C_n^{(k)} \setminus S_n^{(k)}$,

$$|c| \geq \max_{s \in S_n^{(k)}} |s|. \quad (2)$$

Since by assumption $C_n^{(m)} = C_n \in \mathbb{O}_n$, we must have $k < m$. We will finish the proof by showing that regardless of the value of k , there is a way to effectively increase it by one. More precisely, we establish the following result:

Lemma 6: For the codes $S_n^{(k)}$ and C_n defined above, there is an integer $d \leq m - k$ and codes $\hat{S}_n^{(k+1)}, \hat{S}_n^{(k+2)}, \dots, \hat{S}_n^{(k+d)} \in \mathbb{S}_n$ for which $S_n^{(k)} \rightarrow \hat{S}_n^{(k+1)} \rightarrow \hat{S}_n^{(k+2)} \rightarrow \dots \rightarrow \hat{S}_n^{(k+d)} = C_n$.

Proof: By assumption, $S_n^{(k)} \neq C_n$. For $i \in \{k, k+1, \dots, m\}$, define

$$F^{(i)} = \{\sigma \in C_n^{(i)} : \sigma \text{ has a prefix in } S_n^{(k)}\} \quad (3)$$

$$\text{and } G^{(i)} = \{\sigma \in C_n^{(i)} : \sigma \text{ has no prefix in } S_n^{(k)}\}. \quad (4)$$

The sets $F^{(i)}$ and $G^{(i)}$ are clearly disjoint, and

$$C_n^{(i)} = F^{(i)} \cup G^{(i)}. \quad (5)$$

For $i \geq k$, each w_i defined by (1) satisfies $w_i \in F^{(i)}$ or $w_i \in G^{(i)}$, but not both. Consider the case where $w_i \in G^{(i)}$, $w_i \notin F^{(i)}$. By (1) and (5),

$$F^{(i+1)} \subseteq F^{(i)} \cup ((G^{(i)} \setminus \{w_i\}) \cup \mathcal{N}_n(w_i)). \quad (6)$$

By the argument used in the proof of Theorem 4, every element of $G^{(i)}$ has a prefix in $C_n^{(k)}$. Therefore the definition of $G^{(i)}$ implies that each of its elements, including w_i , has a prefix in $C_n^{(k)} \setminus S_n^{(k)}$. Hence every element of the sets $\mathcal{N}_n(w_i)$ and $(G^{(i)} \setminus \{w_i\}) \cup \mathcal{N}_n(w_i)$ has a prefix in $C_n^{(k)} \setminus S_n^{(k)}$. To arrive at a contradiction, suppose $v \in (G^{(i)} \setminus \{w_i\}) \cup \mathcal{N}_n(w_i)$ has a prefix in $S_n^{(k)}$, say s . Let c be the prefix of v in $C_n^{(k)} \setminus S_n^{(k)}$. Since both s and c are prefixes of v , either s is a prefix of c or c is a prefix of s . Observe that $s, c \in C_n^{(k)} \cup S_n^{(k)}$, and so $C_n^{(k)} \cup S_n^{(k)}$ does not satisfy the prefix condition. However, $C_n^{(k)} \cup S_n^{(k)}$ is a symmetric fix-free code because the rules for constructing $C_n^{(k)}$ and $S_n^{(k)}$ imply that

$$C_n^{(k)} \cup S_n^{(k)} \subseteq (C_n^{(k-1)} \setminus \{w_{k-1}\}) \cup \mathcal{N}_n(w_{k-1}),$$

and the right-hand side of the preceding relation describes a symmetric fix-free code. This contradiction implies that no element of $(G^{(i)} \setminus \{w_i\}) \cup \mathcal{N}_n(w_i)$ has a prefix in $S_n^{(k)}$. Therefore, we find from (6) that

$$F^{(i+1)} \cap ((G^{(i)} \setminus \{w_i\}) \cup \mathcal{N}_n(w_i)) = \emptyset. \quad (7)$$

Therefore (6) and (7) imply that for $i \geq k$,

$$F^{(i+1)} \subseteq F^{(i)} \text{ if } w_i \notin F^{(i)}. \quad (8)$$

In the derivation of (7) we argued that $C_n^{(k)} \cup S_n^{(k)}$ is a symmetric fix-free code and hence satisfies the prefix condition. Observe that $C_n^{(k)} \cup S_n^{(k)} = (C_n^{(k)} \setminus S_n^{(k)}) \cup S_n^{(k)}$. Therefore no element of $C_n^{(k)} \setminus S_n^{(k)}$ has a prefix in $S_n^{(k)}$, or equivalently,

$$C_n^{(k)} \setminus S_n^{(k)} \subseteq G^{(k)}. \quad (9)$$

Since every element of $C_n^{(k)} \cap S_n^{(k)}$ has a prefix in $S_n^{(k)}$, it follows that

$$C_n^{(k)} \cap S_n^{(k)} \subseteq F^{(k)}. \quad (10)$$

By (5), we have $F^{(k)} \cup G^{(k)} = C_n^{(k)} = (C_n^{(k)} \cap S_n^{(k)}) \cup (C_n^{(k)} \setminus S_n^{(k)})$. Therefore, (9) and (10) imply that $F^{(k)} = C_n^{(k)} \cap S_n^{(k)}$, and so

$$F^{(k)} \subseteq S_n^{(k)}. \quad (11)$$

To continue our argument, we will next show that

$$F^{(m)} = C_n \text{ and } G^{(m)} = \emptyset. \quad (12)$$

To arrive at a contradiction, assume $v \in G^{(m)}$. Then there is a string $s \in S_n^{(k)}$ which is not the prefix of any codeword of C_n . By Theorem 4, v has a prefix in $C_n^{(k)}$, say c . Since $v \in G^{(m)}$, it follows that $c \in C_n^{(k)} \setminus S_n^{(k)}$. By (2) we have $|v| \geq |c| \geq |s|$. There are two cases to consider:

- 1) $|v| > |s|$: Since s is a palindrome which is not the prefix of any codeword in C_n , we have that $(C_n \setminus \{v\}) \cup \{s\}$ is a symmetric fix-free code with n codewords which is better than C_n for any probabilistic source. Hence, $C_n \notin \mathbb{O}_n$, which contradicts our assumption.
- 2) $|v| = |s|$: Then $v = c$ and so $v \in C_n^{(k-1)} \cup \mathcal{N}_n(w_{k-1}) \setminus \{w_{k-1}\}$ and $v \notin S_n^{(k)}$. Therefore $(S_n^{(k)} \setminus \{s\}) \cup \{v\}$ has the same length sequence as $S_n^{(k)}$ and greater overlap with C_n , which contradicts our assumption about the choice of $S_n^{(k)}$.

We next show that $w_i \in F^{(i)}$ for some $i \in \{k, k+1, \dots, m-1\}$. Suppose that $w_i \notin F^{(i)}$ for all $i \geq k$. Then by (8) and (11),

$$F^{(m)} \subseteq \dots \subseteq F^{(k)} \subseteq S_n^{(k)}. \quad (13)$$

By (12), (13), and the fact that $C_n, S_n^{(k)} \in \mathbb{S}_n$, we obtain $C_n = S_n^{(k)}$, which contradicts our assumption.

Define the set $\{i_k, \dots, i_{k+d-1}\} \subseteq \{k, \dots, m-1\}$ to be the collection of indices for which $w_{i_l} \in F^{(i_l)}$, $l \in \{k, \dots, k+d-1\}$ and $w_i \in G^{(i)}$, $i \notin \{i_k, \dots, i_{k+d-1}\}$. Then by (8) and (11), we obtain

$$F^{(i_k)} \subseteq \dots \subseteq F^{(k)} \subseteq S_n^{(k)} \quad (14)$$

$$F^{(i_{l+1})} \subseteq \dots \subseteq F^{(i_l+1)}, \quad l \in \{k, \dots, k+d-2\} \quad (15)$$

Since $C_n^{(i_l+1)} = F^{(i_l+1)} \cup G^{(i_l+1)} \subseteq ((F^{(i_l)} \setminus \{w_{i_l}\}) \cup \mathcal{N}_n(w_{i_l})) \cup G^{(i_l)}$ and $w_{i_l} \in F^{(i_l)}$ implies that every element of $\mathcal{N}_n(w_{i_l})$ has a prefix in $S_n^{(k)}$, we find that

$$F^{(i_{l+1})} \subseteq (F^{(i_l)} \setminus \{w_{i_l}\}) \cup \mathcal{N}_n(w_{i_l}), \quad l \in \{k, \dots, k+d-2\}. \quad (16)$$

From (14), we obtain $w_{i_k} \in F^{(i_k)} \subseteq S_n^{(k)}$. By (14) and (16), we can verify that

$$F^{(i_{k+1})} \subseteq (S_n^{(k)} \setminus \{w_{i_k}\}) \cup \mathcal{N}_n(w_{i_k}).$$

Therefore, there exists a symmetric fix-free code $\hat{S}_n^{(k+1)} \in \mathbb{S}_n$ such that

$$F^{(i_{k+1})} \subseteq \hat{S}_n^{(k+1)} \subseteq (S_n^{(k)} \setminus \{w_{i_k}\}) \cup \mathcal{N}_n(w_{i_k}), \quad (17)$$

and so $S_n^{(k)} \rightarrow \hat{S}_n^{(k+1)}$. Similarly, we can construct a sequence of symmetric fix-free codes $\hat{S}_n^{(k+2)}, \dots, \hat{S}_n^{(k+d)} \in \mathbb{S}_n$ for which

$$F^{(i_l+1)} \subseteq \hat{S}_n^{(l+1)} \subseteq (\hat{S}_n^{(l)} \setminus \{w_{i_l}\}) \cup \mathcal{N}_n(w_{i_l}), \quad l \in \{k+1, \dots, k+d-1\}. \quad (18)$$

Hence, $S_n^{(k)} \rightarrow \hat{S}_n^{(k+1)} \rightarrow \dots \rightarrow \hat{S}_n^{(k+d)}$.

By (12), (15), and (18), we can show that $C_n = F^{(m)} \subseteq \hat{S}_n^{(k+d)}$. Because $C_n, \hat{S}_n^{(k+d)} \in \mathbb{S}_n$, we have $C_n = \hat{S}_n^{(k+d)}$. Thus,

$$S_n^{(k)} \rightarrow \hat{S}_n^{(k+1)} \rightarrow \hat{S}_n^{(k+2)} \rightarrow \dots \rightarrow \hat{S}_n^{(k+d)} = C_n$$

with $k+d \leq m$. □

To reiterate the result, if $k-1 \neq m$ we can alter the generation of code C_n from

$$\begin{aligned} R_n = C_n^{(0)} &\Rightarrow \dots \Rightarrow C_n^{(k-1)} \rightarrow C_n^{(k)} \rightarrow \dots \rightarrow C_n^{(m)} = C_n \\ \text{to } R_n = C_n^{(0)} &\Rightarrow \dots \Rightarrow C_n^{(k-1)} \Rightarrow S_n^{(k)} \rightarrow \dots \rightarrow \hat{S}_n^{(k+d)} = C_n \end{aligned}$$

for some $k+d \leq m$. By repeatedly applying this argument we obtain the result. □

Comment: There is some evidence that for codes in \mathbb{O}_n the number m of \Rightarrow operations needed is $O(n \log_2 n)$. In [13, Prop. 2.6] we showed that the average number of bits per symbol of the optimal symmetric fix-free code is at most $2\mathcal{H} + 1$, where \mathcal{H} is the binary entropy of the source. Suppose the source probabilities are $p_1 \geq p_2 \geq \dots \geq p_n$. Then $m \leq \sum_{i=1}^n (l_i - 1) \leq \sum_{i=1}^n p_i (l_i - 1) / p_n \leq 2\mathcal{H} / p_n$.

4. Simplifying the Search for Optimal Symmetric Fix-Free Codes

The sequence of symmetric fix-free codes from the root code R_n to an optimal code $C_n \in \mathbb{O}_n$ as defined in Theorem 5 is often not unique. The following result further specifies such codes.

Lemma 7: For any code $C_n \in \mathbb{O}_n$, suppose $R_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_m} S_n^{(m)} = C_n$. Then this is a shortest sequence of symmetric fix-free codes transforming R_n to C_n via repeated uses of the \Rightarrow operation if and only if π_i is a prefix of at least one codeword in C_n for each $i \in \{1, \dots, m\}$.

Proof: Let us first consider the case where the condition is not satisfied. Let $l \in \{1, \dots, m\}$ denote the maximum index for which π_l is not a prefix of any codeword in C_n . Observe that it is impossible to have $l = m$ because $C_n = S_n^{(m)}$ has a nonempty intersection with $\mathcal{N}_n(\pi_m)$ since C_n and $S_n^{(m-1)}$ both have n codewords. Therefore, $l < m$. For $i \geq l+1$, π_i is a prefix of at least one codeword in C_n , so π_l cannot be a prefix of π_i . Thus, by the definition of the \Rightarrow operation we can write

$$S_n^{(i)} \setminus \mathcal{N}_n(\pi_l) \subseteq (S_n^{(i-1)} \setminus \mathcal{N}_n(\pi_l)) \cup \mathcal{N}_n(\pi_i) \setminus \{\pi_i\}, \quad i \in \{l+1, \dots, m\}. \quad (19)$$

Since π_l is not a prefix of π_i , $i \in \{l+1, \dots, m\}$, it follows from (19) that for $i \geq l+1$,

$$\pi_i \in S_n^{(i-1)} \setminus \mathcal{N}_n(\pi_l). \quad (20)$$

We will use induction to establish the existence of codes $C_n^{(l+1)}, \dots, C_n^{(m)} = C_n \in \mathbb{S}_n$ satisfying

$$S_n^{(i)} \setminus \mathcal{N}_n(\pi_l) \subseteq C_n^{(i)}, \quad i \in \{l+1, \dots, m\}, \quad (21)$$

$$\text{and } S_n^{(l-1)} \xrightarrow{\pi_{l+1}} C_n^{(l+1)} \xrightarrow{\pi_{l+2}} C_n^{(l+2)} \dots \xrightarrow{\pi_m} C_n^{(m)} = C_n. \quad (22)$$

For the basis step, the definition of the \Rightarrow operation implies

$$S_n^{(l)} \setminus \mathcal{N}_n(\pi_l) \subseteq S_n^{(l-1)} \setminus \{\pi_l\}. \quad (23)$$

Furthermore, we have seen that π_l is not a prefix of π_{l+1} . By (20) and (23) we have

$$\pi_{l+1} \in S_n^{(l-1)}. \quad (24)$$

It follows from (19) that

$$S_n^{(l+1)} \setminus \mathcal{N}_n(\pi_l) \subseteq (S_n^{(l)} \setminus \mathcal{N}_n(\pi_l)) \cup \mathcal{N}_n(\pi_{l+1}) \setminus \{\pi_{l+1}\} \subseteq S_n^{(l)} \cup \mathcal{N}_n(\pi_{l+1}) \setminus \{\pi_{l+1}\}. \quad (25)$$

Observe that $S_n^{(l+1)} \setminus \mathcal{N}_n(\pi_l)$ contains at most n words and $S_n^{(l)} \cup \mathcal{N}_n(\pi_{l+1}) \setminus \{\pi_{l+1}\}$ contains at least n words. Therefore, by (24) and (25), there exists $C_n^{(l+1)} \in \mathbb{S}_n$ such that $S_n^{(l+1)} \setminus \mathcal{N}_n(\pi_l) \subseteq C_n^{(l+1)}$ and $S_n^{(l-1)} \xrightarrow{\pi_{l+1}} C_n^{(l+1)} \xrightarrow{\pi_{l+2}}$

For the inductive step, suppose that for some $l+1 \leq k < m$ we have found symmetric fix-free codes $C_n^{(l+1)}, \dots, C_n^{(k)} \in \mathbb{S}_n$ which satisfy (21) and $S_n^{(l-1)} \xrightarrow{\pi_{l+1}} C_n^{(l+1)} \xrightarrow{\pi_{l+2}} C_n^{(l+2)} \dots \xrightarrow{\pi_k} C_n^{(k)}$. We next generate $C_n^{(k+1)}$. By (20), (21), and (19) we have

$$\begin{aligned} \pi_{k+1} &\in S_n^{(k)} \setminus \mathcal{N}_n(\pi_l) \subseteq C_n^{(k)} \text{ and} \\ S_n^{(k+1)} \setminus \mathcal{N}_n(\pi_l) &\subseteq (S_n^{(k)} \setminus \mathcal{N}_n(\pi_l)) \cup \mathcal{N}_n(\pi_{k+1}) \setminus \{\pi_{k+1}\} \subseteq C_n^{(k)} \cup \mathcal{N}_n(\pi_{k+1}) \setminus \{\pi_{k+1}\}. \end{aligned}$$

Like the argument for the basis step, there exists $C_n^{(k+1)} \in \mathbb{S}_n$ for which $S_n^{(k+1)} \setminus \mathcal{N}_n(\pi_l) \subseteq C_n^{(k+1)}$ and $S_n^{(l-1)} \xrightarrow{\pi_{l+1}} C_n^{(l+1)} \xrightarrow{\pi_{l+2}} C_n^{(l+2)} \dots \xrightarrow{\pi_{k+1}} C_n^{(k+1)}$. At $k+1 = m$ we have $C_n = S_n^{(m)} \setminus \mathcal{N}_n(\pi_l)$, and therefore $C_n = S_n^{(m)} = C_n^{(m)}$.

We have established a sequence of symmetric fix-free codes $S_n^{(0)}, S_n^{(1)}, \dots, S_n^{(l-1)}, C_n^{(l+1)}, \dots, C_n^{(m)}$ for which $R_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_{l-1}} S_n^{(l-1)} \xRightarrow{\pi_{l+1}} C_n^{(l+1)} \xRightarrow{\pi_{l+2}} \dots \xRightarrow{\pi_m} C_n^{(m)} = C_n$. By the argument used in the proof of Theorem 5, these relations imply the existence a sequence of symmetric fix-free codes $S_n^{(0)}, D_n^{(1)}, D_n^{(2)}, \dots, D_n^{(j)} = C_n \in \mathbb{S}_n$ with $j \leq m-1$ for which $R_n = S_n^{(0)} \Rightarrow D_n^{(1)} \Rightarrow D_n^{(2)} \Rightarrow \dots \Rightarrow D_n^{(j)} = C_n$, which demonstrates that $R_n = S_n^{(0)}, S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)} = C_n$ is not a shortest sequence of codes transforming R_n to C_n via repeated uses of the \Rightarrow operation.

For the converse, given an arbitrary code $C_n \in \mathbb{S}_n$ let C^{prefix} be the set of palindromes (not including 1) which are proper prefixes of at least one codeword in C_n . Suppose we are given a set of codes $S_n^{(0)}, S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)} \in \mathbb{S}_n$ and palindromes $\{\pi_1, \dots, \pi_m\}$ defined by $R_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_m} S_n^{(m)} = C_n$. We will show that $C^{\text{prefix}} \subseteq \{\pi_1, \dots, \pi_m\}$.

For each $w \in C_n$, define $C^{\text{prefix}}(w)$ to be the set of palindromes (not including 1) which are proper prefixes of w . Then $C^{\text{prefix}} = \cup_{w \in C_n} C^{\text{prefix}}(w)$. If $w \in R_n$, then $C^{\text{prefix}}(w) = \emptyset$. Otherwise, there is an ordering of the $\eta_w \geq 1$ strings in $C^{\text{prefix}}(w)$, say $\sigma_w^{(1)}, \dots, \sigma_w^{(\eta_w)}$, so that $\sigma_w^{(1)} \in R_n$, $\sigma_w^{(i+1)} \in \mathcal{N}_n(\sigma_w^{(i)})$ for $i \in \{1, \dots, \eta_w - 1\}$, and $w \in \mathcal{N}_n(\sigma_w^{(\eta_w)})$. Observe that $w \in C_n$ implies that $\sigma_w^{(i)} \in \{\pi_1, \dots, \pi_m\}$ for all $w \notin R_n$ and $i \in \{1, \dots, \eta_w\}$. Therefore $C^{\text{prefix}}(w) \subseteq \{\pi_1, \dots, \pi_m\}$ for all $w \in C_n$, and so

$$C^{\text{prefix}} \subseteq \{\pi_1, \dots, \pi_m\}. \quad (26)$$

Because C^{prefix} is determined only by C_n , in order for $S_n^{(0)}, S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(m)} \in \mathbb{S}_n$ to be a shortest sequence of codes transforming R_n to C_n via uses of the \Rightarrow operation, it suffices to show that

$$C^{\text{prefix}} = \{\pi_1, \dots, \pi_m\}. \quad (27)$$

The assumption $\{\pi_1, \dots, \pi_m\} \subseteq C^{\text{prefix}}$ together with (26) results in (27). \square

Given Lemma 7 and (27), we next show

Theorem 8: For any code $C_n \in \mathbb{O}_n$, suppose $R_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_m} S_n^{(m)} = C_n$ is a shortest sequence of codes in \mathbb{S}_n transforming R_n to C_n via uses of the \Rightarrow operation. Define $C^{\text{prefix}} = \{\pi_1, \dots, \pi_m\}$. Then any ordering $\sigma_1, \sigma_2, \dots, \sigma_m$ of the elements of C^{prefix} with $i < j$ whenever σ_i is a prefix of σ_j corresponds to a sequence of symmetric fix-free codes $C_n^{(\Sigma,0)}, C_n^{(\Sigma,1)}, C_n^{(\Sigma,2)}, \dots, C_n^{(\Sigma,m)} \in \mathbb{S}_n$ satisfying $R_n = C_n^{(\Sigma,0)} \xRightarrow{\sigma_1} C_n^{(\Sigma,1)} \xRightarrow{\sigma_2} C_n^{(\Sigma,2)} \xRightarrow{\sigma_3} \dots \xRightarrow{\sigma_m} C_n^{(\Sigma,m)} = C_n$.

Proof: There are two main parts to the proof. In the first we show that there is a set of transformations starting from $\{\pi_1, \dots, \pi_m\}$ and ending in $\{\sigma_1, \dots, \sigma_m\}$ which at each step involves a transposition of an adjacent pair of strings while maintaining the invariant that any palindrome (not including 1) which is a proper prefix of a palindrome in the list always precedes it. In the second part we consider the effect of a (valid) transposition of an adjacent pair of strings in devising shortest transformation from R_n to $C_n \in \mathbb{O}_n$ via uses of the \Rightarrow operation.

For the first part of the proof, for a sequence (of numbers or strings) $A = (a_1, a_2, \dots, a_m)$, define A_i , $i \in \{1, \dots, m-1\}$, as the permutation of A obtained by transposing a_i and a_{i+1} . For example, if $A = (1, 2, 3, 4)$, then $A_1 = (2, 1, 3, 4)$, $A_2 = (1, 3, 2, 4)$, $A_3 = (1, 2, 4, 3)$. We have the following result.

Lemma 9: For (π_1, \dots, π_m) and $(\sigma_1, \dots, \sigma_m)$ defined in Theorem 8, define $\Omega^0 = (\pi_1, \dots, \pi_m)$. Then there is a number $k < m^2$, a sequence of indices $a_1, \dots, a_k \in \{1, \dots, m-1\}$, and a sequence of pairwise permutations starting from Ω^0 with $\Omega^i = (\Omega^{i-1})_{a_i}$ and $\Omega^k = (\sigma_1, \dots, \sigma_m)$ such that for all i , Ω^i satisfies the constraint that the proper prefixes in the list of each palindrome precede it in the ordering.

Proof: Suppose we know $\Omega^0, \dots, \Omega^i = (w_1^i, \dots, w_m^i)$, and we wish to construct Ω^{i+1} . Let h_i be the maximum index for which $w_g^i \neq \sigma_g$. Then there is some $l_i < h_i$ for which $w_{l_i}^i = \sigma_{h_i}$. We claim that we can choose $\Omega^{i+1} = (\Omega^i)_{l_i}$; i.e., $w_{l_i}^i$ is not a prefix of $w_{l_i+1}^i$. This is clearly true if σ_{h_i} is not a prefix of σ_j , $j \neq h_i$. If $\sigma_{h_i} = w_{l_i}^i$ is a proper prefix of some $\sigma_j = w_{l_i+1}^i$, then by assumption $j > h_i$, and hence h_i is not the maximum index for which $w_g^i \neq \sigma_g$.

Given this choice of Ω^{i+1} , let us consider the ordered pair (l_{i+1}, h_{i+1}) . If $l_i + 1 < h_i$, then $(l_{i+1}, h_{i+1}) = (l_i + 1, h_i)$, and if $l_i + 1 = h_i$, then $h_{i+1} < h_i$. Since $(l_i, h_i) \neq (l_j, h_j)$ for $i \neq j$, eventually the sequence of pairwise permutations will terminate in $\Omega^k = (\sigma_1, \dots, \sigma_m)$. \square

For the second part of the proof of Theorem 8, we are given that for Ω^0 , $R_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_m} S_n^{(m)} = C_n$ is a shortest sequence of codes in \mathbb{S}_n transforming R_n to C_n via uses of the \Rightarrow operation. Next suppose that for some $i \geq 0$, there is a sequence of symmetric fix-free codes $C_n^{(\Omega^i,0)}, C_n^{(\Omega^i,1)}, C_n^{(\Omega^i,2)}, \dots, C_n^{(\Omega^i,m)} \in \mathbb{S}_n$ satisfying $R_n = C_n^{(\Omega^i,0)} \xRightarrow{w_1^i} C_n^{(\Omega^i,1)} \xRightarrow{w_2^i} C_n^{(\Omega^i,2)} \xRightarrow{w_3^i} \dots \xRightarrow{w_m^i} C_n^{(\Omega^i,m)} = C_n$. By Lemma 9, to complete the proof of Theorem 8 it suffices to show that there is a sequence of symmetric fix-free codes $C_n^{(\Omega^{i+1},0)}, C_n^{(\Omega^{i+1},1)}, C_n^{(\Omega^{i+1},2)}, \dots, C_n^{(\Omega^{i+1},m)} \in \mathbb{S}_n$ satisfying $R_n = C_n^{(\Omega^{i+1},0)} \xRightarrow{w_1^{i+1}} C_n^{(\Omega^{i+1},1)} \xRightarrow{w_2^{i+1}} C_n^{(\Omega^{i+1},2)} \xRightarrow{w_3^{i+1}} \dots \xRightarrow{w_m^{i+1}} C_n^{(\Omega^{i+1},m)} = C_n$.

From the proof of Lemma 9, we have the following relationship between $\Omega^{i+1} = (w_1^{i+1}, \dots, w_m^{i+1})$ and $\Omega^i = (w_1^i, \dots, w_m^i)$:

$$w_j^{i+1} = \begin{cases} w_j^i, & j \notin \{l_i, l_{i+1}\} \\ w_{l_i+1}^i, & j = l_i \\ w_{l_i}^i, & j = l_i + 1 \end{cases}$$

In the proof of Lemma 9 we argued that $w_{l_i}^i$ is not a prefix of $w_{l_i+1}^i$ (or vice versa). Therefore, for $j < l_i$ we will choose $C_n^{(\Omega^{i+1}, j)} = C_n^{(\Omega^i, j)}$. If there exists $C_n^{(\Omega^{i+1}, l_i)} \in \mathbb{S}_n$ for which

$$C_n^{(\Omega^i, l_i-1)} \xrightarrow{w_{l_i+1}^i} C_n^{(\Omega^{i+1}, l_i)} \xrightarrow{w_{l_i}^i} C_n^{(\Omega^i, l_i+1)}, \quad (28)$$

then for $j \geq l_i + 1$ we can choose $C_n^{(\Omega^{i+1}, j)} = C_n^{(\Omega^i, j)}$. We next establish the existence of $C_n^{(\Omega^{i+1}, l_i)}$ to satisfy (28). To simplify notation, define

$$S_n = C_n^{(\Omega^i, l_i-1)}, \quad I_n = C_n^{(\Omega^i, l_i)}, \quad S'_n = C_n^{(\Omega^i, l_i+1)}, \quad \omega_1 = w_{l_i}, \quad \omega_2 = w_{l_i+1}$$

so that

$$S_n \xrightarrow{\omega_1} I_n \xrightarrow{\omega_2} S'_n. \quad (29)$$

Let $C_n(\omega_1)$ be the subset of words in C_n which have ω_1 as a prefix. By Lemma 7, $C_n(\omega_1) \neq \emptyset$. Since C_n and S_n both have n strings, there exists $S_n(\omega_1) \subseteq S_n$ with $|S_n(\omega_1)| = |C_n(\omega_1)|$, $\omega_1 \in S_n(\omega_1)$, and $\omega \in S_n(\omega_1)$ is not a prefix of any codeword in C_n if $\omega \neq \omega_1$. Observe that $(C_n \setminus C_n(\omega_1)) \cup S_n(\omega_1) \in \mathbb{S}_n$. To arrive at a contradiction, suppose $\min_{\sigma \in \mathcal{N}_n(\omega_1)} |\sigma| \geq \max_{\sigma \in S_n} |\sigma|$. Then $\min_{\sigma \in C_n(\omega_1)} |\sigma| \geq \max_{\sigma \in S_n(\omega_1)} |\sigma|$ and $\min_{\sigma \in C_n(\omega_1)} |\sigma| \geq |\omega_1| + 1$. Therefore the code $(C_n \setminus C_n(\omega_1)) \cup S_n(\omega_1)$ is a strictly better symmetric fix-free code than C_n for any choice of source probabilities, contradicting the assumption that $C_n \in \mathbb{O}_n$. Hence,

$$\min_{\sigma \in \mathcal{N}_n(\omega_1)} |\sigma| < \max_{\sigma \in S_n} |\sigma|. \quad (30)$$

We likewise have

$$\min_{\sigma \in \mathcal{N}_n(\omega_2)} |\sigma| < \max_{\sigma \in I_n} |\sigma|. \quad (31)$$

Since $\omega_1 \in S_n$ is a prefix of at least one codeword in C_n , it must also be a prefix of at least one codeword in S'_n . Furthermore, because $\omega_1, \omega_2 \in S_n$ and are distinct, ω_1 is not a prefix of any string in $\mathcal{N}_n(\omega_2)$. Hence,

$$\mathcal{N}_n(\omega_1) \cap S'_n \neq \emptyset. \quad (32)$$

In order to continue our discussion of the transposition of a successive pair of \Rightarrow operations, we introduce the following notation:

$$\begin{aligned} I_n &= \hat{I}(\omega_1) \cup \hat{S}(\omega_1) \\ \hat{S}(\omega_1) &\subseteq S_n \setminus \{\omega_1\} \\ \hat{I}(\omega_1) &\subseteq \mathcal{N}_n(\omega_1) \\ S'_n &= \tilde{S}(\omega_1, \omega_2) \cup \hat{J}(\omega_1) \cup \tilde{J}(\omega_2) \\ \tilde{S}(\omega_1, \omega_2) &\subseteq \hat{S}(\omega_1) \setminus \{\omega_2\} \subseteq S_n \setminus \{\omega_1, \omega_2\} \\ \hat{J}(\omega_1) &\subseteq \hat{I}(\omega_1) \subseteq \mathcal{N}_n(\omega_1) \\ \tilde{J}(\omega_2) &\subseteq \mathcal{N}_n(\omega_2) \end{aligned}$$

We have the following result.

Proposition 10: There exists $J_n \in \mathbb{S}_n$ such that $\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2) \subseteq J_n$ and $S_n \xrightarrow{\omega_2} J_n$.

Proof: By (29), $I_n \xrightarrow{\omega_2} S'_n$, and it follows that $\mathcal{N}_n(\omega_2) \neq \emptyset$. Therefore there is at least one choice for $I'_n \in \mathbb{S}_n$ for which

$$S_n \xrightarrow{\omega_2} I'_n. \quad (33)$$

We will next show that $\omega_1 \in I'_n$. To arrive at a contradiction, suppose $\omega_1 \notin I'_n$. Then by the definition of the \Rightarrow operation

$$|\omega_1| \geq \max_{\sigma \in I'_n} |\sigma|. \quad (34)$$

Define sets $S^*(\omega_2)$ and $J^*(\omega_2)$ by

$$\begin{aligned} I'_n &= S^*(\omega_2) \cup J^*(\omega_2) \\ S^*(\omega_2) &\subseteq S_n \setminus \{\omega_2\} \\ J^*(\omega_2) &\subseteq \mathcal{N}_n(\omega_2) \end{aligned}$$

Since $S^*(\omega_2) \subseteq I'_n$, (34) implies

$$|\omega_1| \geq \max_{\sigma \in S^*(\omega_2)} |\sigma|. \quad (35)$$

The relation $S_n \xrightarrow{\omega_1} I_n$ implies that I_n contains all elements of S_n with length at most $|\omega_1|$, and combined with (35) we obtain $S^*(\omega_2) \subseteq I_n \setminus \{\omega_2\}$. Thus,

$$I'_n = S^*(\omega_2) \cup J^*(\omega_2) \subseteq (I_n \setminus \{\omega_2\}) \cup \mathcal{N}_n(\omega_2).$$

The previous relation and (29) imply

$$I_n \xrightarrow{\omega_2} I'_n \text{ and } I_n \xrightarrow{\omega_2} S'_n. \quad (36)$$

Thus, the difference between the \rightarrow and \Rightarrow operations, (34), (36), and (32) imply

$$|\omega_1| \geq \max_{\sigma \in I'_n} |\sigma| \geq \max_{\sigma \in S'_n} |\sigma| \geq \min_{\sigma \in \mathcal{N}_n(\omega_1)} |\sigma| > |\omega_1|,$$

which is impossible. Hence the assumption that $\omega_1 \notin I'_n$ was false. Therefore

$$S_n \xrightarrow{\omega_2} I'_n \text{ implies } \omega_1 \in I'_n. \quad (37)$$

Recall that $S'_n = \tilde{S}(\omega_1, \omega_2) \cup \hat{J}(\omega_1) \cup \tilde{J}(\omega_2)$. By (32) we have $\hat{J}(\omega_1) \neq \emptyset$. Since S'_n has n codewords, it follows that $\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2)$ has at most n elements. To arrive at a contradiction, suppose there is no J_n that simultaneously satisfies $S_n \xrightarrow{\omega_2} J_n$ and $\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2) \subseteq J_n$. Then choose some set J_n for which $S_n \xrightarrow{\omega_2} J_n$. Since $\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2) \not\subseteq J_n$, the relation $J_n \subseteq S_n \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$ and the definition of the \Rightarrow operation imply the existence of $x \in J_n \setminus (\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2))$ and $y \in \tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2) \setminus J_n$ with $|y| > |x|$. By (37) we know $\omega_1 \in J_n$, so $x \neq \omega_1$ and $y \neq \omega_1$. Therefore, $y \in \tilde{S}(\omega_1, \omega_2) \cup \tilde{J}(\omega_2)$; i.e., $y \in S'_n$. Similarly, $x \in J_n \subseteq S_n \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$ and $x \notin (\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2))$ implies that $x \notin S'_n$. Since $x \in J_n$ and $x \neq \omega_1$ we consider two exhaustive cases for the membership of x :

- $x \in \hat{S}(\omega_1) \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$: Since $\hat{S}(\omega_1) \subseteq I_n$ we have $x \in I_n \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$. Thus, there exists $S_n'' \in \mathbb{S}_n$ such that $x \in S_n''$ and $I_n \xrightarrow{\omega_2} S_n''$. Recall that $I_n \xrightarrow{\omega_2} S_n'$. We saw earlier that $x \notin S_n'$ and $y \in S_n'$. Therefore, $|x| \geq \max_{\sigma \in S_n'} |\sigma| \geq |y|$, which violates our earlier argument that $|y| > |x|$.
- $x \in S_n \setminus \{\hat{S}(\omega_1) \cup \{\omega_1\}\}$: Since $x \in S_n \setminus \{\omega_1\} \subseteq S_n \cup \mathcal{N}_n(\omega_1) \setminus \{\omega_1\}$, there exists $I_n'' \in \mathbb{S}_n$ such that $x \in I_n''$ and $S_n \xrightarrow{\omega_1} I_n''$. Since $S_n \cap \mathcal{N}_n(\omega_1) = \emptyset$, we have $x \notin \mathcal{N}_n(\omega_1)$. We also assume $x \notin \hat{S}(\omega_1)$. It follows that $x \notin I_n$. Recall that $S_n \xrightarrow{\omega_1} I_n$. Therefore, $|x| \geq \max_{\sigma \in I_n} |\sigma|$. By (31), $\max_{\sigma \in I_n} |\sigma| > \min_{\sigma \in \mathcal{N}_n(\omega_2)} |\sigma|$. S_n' consists of the smallest elements of $I_n \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$, so $\max_{\sigma \in I_n} |\sigma| \geq \max_{\sigma \in S_n'} |\sigma|$. We have already seen that $y \in S_n'$. Combining these observations we obtain $|x| \geq \max_{\sigma \in I_n} |\sigma| \geq \max_{\sigma \in S_n'} |\sigma| \geq |y|$, which violates our earlier argument that $|y| > |x|$.

Therefore, our assumption was false, and this establishes the proposition. \square

Proposition 11: For the symmetric fix-free code J_n described by Proposition 10,

$$J_n \xrightarrow{\omega_1} S_n'.$$

Proof: Recall that $S_n' = \tilde{S}(\omega_1, \omega_2) \cup \hat{J}(\omega_1) \cup \tilde{J}(\omega_2)$ and $\tilde{S}(\omega_1, \omega_2) \cup \{\omega_1\} \cup \tilde{J}(\omega_2) \subseteq J_n$. Thus, $S_n' \subseteq J_n \cup \mathcal{N}_n(\omega_1) \setminus \{\omega_1\}$. Therefore, $J_n \xrightarrow{\omega_1} S_n'$. To arrive at a contradiction, suppose $J_n \not\xrightarrow{\omega_1} S_n'$. Then choose some S_n'' to satisfy $J_n \xrightarrow{\omega_1} S_n''$. There exists $x \in S_n'' \setminus S_n'$ and $y \in S_n' \setminus S_n''$ such that $|x| < |y|$. Observe that $x \in J_n \cup \mathcal{N}_n(\omega_1) \setminus \{\omega_1\} \subseteq S_n \cup \mathcal{N}_n(\omega_1) \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_1, \omega_2\}$. There are two exhaustive cases for the membership of x :

- $x \in I_n \cup \mathcal{N}_n(\omega_2) \setminus \{\omega_2\}$: There exists $\tilde{S}_n' \in \mathbb{S}_n$ with $x \in \tilde{S}_n' \setminus S_n'$ and $I_n \xrightarrow{\omega_2} \tilde{S}_n'$. By (29), $I_n \xrightarrow{\omega_2} S_n'$. Since $y \in S_n'$ it follows that $|x| \geq \max_{\sigma \in S_n'} |\sigma| \geq |y|$, which contradicts our assumption that $|x| < |y|$.
- $x \in S_n \cup \mathcal{N}_n(\omega_1) \setminus (I_n \cup \{\omega_1\})$: Since $x \in S_n \cup \mathcal{N}_n(\omega_1) \setminus \{\omega_1\}$, there exists $I_n'' \in \mathbb{S}_n$ such that $x \in I_n'' \setminus I_n$ and $S_n \xrightarrow{\omega_1} I_n''$. By (29), $S_n \xrightarrow{\omega_1} I_n$. Since $x \notin I_n$ we can conclude that $|x| \geq \max_{\sigma \in I_n} |\sigma|$ and repeat the end of the argument for Proposition 10 to obtain a contradiction.

Since our assumption that $J_n \not\xrightarrow{\omega_1} S_n'$ was false, we have established the proposition. \square

To complete the proof of Theorem 8 we choose $C_n^{(\Omega^{i+1}, l_i)} = J_n$. \square

Remark: Lemma 7 and Theorem 8 are important to reduce the computational complexity of the search for optimal codes because by allowing a natural ordering to be imposed on the strings in C^{prefix} one can potentially have a large reduction in the number of sequences of transformations that need to be considered.

Thus far we have provided a way to generate any code in \mathbb{O}_n , but the procedure will also generate some codes in $\mathbb{S}_n \setminus \mathbb{O}_n$. Therefore, it is desirable to provide simple tests to reduce the number of candidate for codes in \mathbb{O}_n . We begin by describing a previously known property of optimal sorted and nondecreasing sequences of codeword lengths corresponding to symmetric fix-free codes. We then offer simplifications of this result, including a generalization of Theorem 2.

Lemma 12: [13, Lemma 2.1] Let (l_1, l_2, \dots, l_n) be the sorted and non-decreasing sequence of codeword lengths corresponding to a symmetric fix-free code and $(l'_1, l'_2, \dots, l'_n)$ be a non-decreasing sequence of natural numbers for which

$$\sum_{j=1}^i l'_j \geq \sum_{j=1}^i l_j \quad \text{for each } i \in \{1, \dots, n\}.$$

Then $(l'_1, l'_2, \dots, l'_n)$ need not be considered as the potential codeword lengths of an optimal symmetric fix-free code.

In the previous result we say length sequence (l_1, l_2, \dots, l_n) *dominates* the sequence $(l'_1, l'_2, \dots, l'_n)$. Let $\mathbb{D}_n \subset \mathbb{S}_n$ be the set of symmetric fix-free codes with sorted and non-decreasing codeword lengths sequences each of which is not dominated by the sorted and non-decreasing codeword length sequence of any other code in \mathbb{S}_n . We have $\mathbb{O}_n \subseteq \mathbb{D}_n$, but it is unknown if $\mathbb{O}_n = \mathbb{D}_n$ for all n .

For symmetric fix-free codes related by the \Rightarrow operation, the n inequalities of Lemma 12 can be reduced to one. We begin with a special case of this result.

Proposition 13: Suppose that the code S'_n is a candidate for membership in \mathbb{O}_n , and let $S_n \in \mathbb{S}_n$ be a code in a shortest transformation from R_n to S'_n through a sequence of \Rightarrow operations. Let (l_1, l_2, \dots, l_n) and $(l'_1, l'_2, \dots, l'_n)$ be the sorted and non-decreasing sequences of codeword lengths of S_n and S'_n , respectively. Suppose that $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$. If the portion of the shortest transformation from S_n to S'_n satisfies either

- $S_n \xRightarrow{\pi_1} S'_n$ or
- there is a sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(h)} \in \mathbb{S}_n$ for some $h \geq 2$ with

$$S_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_h} S_n^{(h)} = S'_n$$

and with π_1 being a prefix of π_i for $i \geq 2$,

then $S'_n \notin \mathbb{O}_n$.

Proof: We begin by considering the first case and later show how to extend the argument to the second case.

We are given that $S'_n \subseteq S_n \cup \mathcal{N}_n(\pi_1) \setminus \{\pi_1\}$. For integers λ let \tilde{S}_n^λ denote the subset of S_n with string lengths greater than λ . By the definition of the \Rightarrow operator, there is some λ for which

$$S'_n \subseteq S_n \cup \mathcal{N}_\lambda(\pi_1) \setminus \{\{\pi_1\} \cup \tilde{S}_n^\lambda\}. \quad (38)$$

Let

$$\begin{aligned} D &= S_n \setminus S'_n \\ D' &= S'_n \setminus S_n \\ m &= |D| = |D'|. \end{aligned}$$

Let (d_1, \dots, d_m) and (d'_1, \dots, d'_m) respectively denote the sorted and non-decreasing sequences of codeword lengths of D and D' . Then

$$|\pi_1| = d_1 < \lambda + 1 \leq d_2 \leq \dots \leq d_m \quad (39)$$

$$d_1 + 1 \leq d'_1 \leq d'_2 \leq \dots \leq d'_m \leq \lambda. \quad (40)$$

The condition $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$ is equivalent to

$$d'_1 - d_1 \geq (d_2 - d'_2) + \dots + (d_m - d'_m), \quad (41)$$

and (39) and (40) imply that

$$d_j \geq d'_j + 1, \quad j \in \{2, \dots, m\}. \quad (42)$$

We would like to show that $\sum_{j=1}^k l'_j \geq \sum_{j=1}^k l_j$, $k \in \{1, 2, \dots, n\}$. Let i be the largest index for which $l_i = d_1$. Then the preceding inequality is an equality for $1 \leq k \leq i-1$. Let ι be the index for which $l_\iota \leq d'_1 < l_{\iota+1}$. Then for $i \leq k \leq \iota-1$,

$$\sum_{j=1}^k l'_j = \sum_{j=1}^{i-1} l_j + \sum_{j=i}^k l_{j+1} \geq \sum_{j=1}^k l_j.$$

For $\iota \leq k \leq n$, suppose that l'_1, l'_2, \dots, l'_k incorporates the g_k shortest new codeword lengths $d'_1, d'_2, \dots, d'_{g_k}$. If $g_k = 1$, then (40) implies $\sum_{j=1}^k (l'_j - l_j) = d'_1 - d_1 \geq 1$. For $2 \leq g_k \leq m$, (41) and (42) imply

$$\sum_{j=1}^k (l'_j - l_j) = d'_1 - d_1 - \sum_{j=2}^{g_k} (d_j - d'_j) \geq 0,$$

as desired.

For the second case, we let $\mathcal{N}_n^*(\sigma)$ denotes the set of all palindromes of length at most n with σ as a proper prefix. The only change needed to the previous discussion is to replace (38) with

$$S'_n \subseteq S_n \cup \mathcal{N}_{\lambda^*}^*(\pi_1) \setminus \{\{\pi_1\} \cup \tilde{S}_n^{\lambda^*}\}$$

for some λ^* and to replace λ with λ^* in (39) and (40). The rest of the proof remains the same as in the first case. \square

We next extend Proposition 13 and simultaneously generalize Theorem 2.

Theorem 14: Consider a code $S'_n \in \mathbb{O}_n$, and suppose $S_n \in \mathbb{S}_n$ is one of the codes in a shortest transformation from R_n to S'_n through a sequence of \Rightarrow operations. Suppose the portion of this shortest transformation from S_n to S'_n involves the sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(h)} \in \mathbb{S}_n$ for some $h \geq 1$ and satisfies

$$S_n = S_n^{(0)} \xrightarrow{\sigma_1} S_n^{(1)} \xrightarrow{\sigma_2} S_n^{(2)} \xrightarrow{\sigma_3} \dots \xrightarrow{\sigma_h} S_n^{(h)} = S'_n.$$

Let (l_1, l_2, \dots, l_n) and $(l'_1, l'_2, \dots, l'_n)$ be the sorted and non-decreasing sequences of codeword lengths of S_n and S'_n , respectively. Let $\tilde{l}_n^{(i)}, i \in \{0, 1, \dots, h\}$, denote the maximum codeword length of $S_n^{(i)}$. Then $l'_n = \tilde{l}_n^{(h)} \leq \tilde{l}_n^{(h-1)} \leq \dots \leq \tilde{l}_n^{(1)} \leq \tilde{l}_n^{(0)} = l_n$ and $\sum_{j=1}^n l'_j < \sum_{j=1}^n l_j$.

Proof: Let $S'_n(\sigma_i)$ be the subset of words in S'_n which have σ_i as a prefix. By Lemma 7, $S'_n(\sigma_i) \neq \emptyset$. Since S'_n and $S_n^{(i-1)}$ both have n strings, there exists $S_n^{(i-1)}(\sigma_i) \subseteq S_n^{(i-1)}$ with $|S_n^{(i-1)}(\sigma_i)| = |S'_n(\sigma_i)|$, $\sigma_i \in S_n^{(i-1)}(\sigma_i)$, and $\sigma \in S_n^{(i-1)}(\sigma_i)$ is not a prefix of any codeword in S'_n if $\sigma \neq \sigma_i$. Observe that $(S'_n \setminus S'_n(\sigma_i)) \cup S_n^{(i-1)}(\sigma_i) \in \mathbb{S}_n$. Observe that if $\min_{\sigma \in \mathcal{N}_n(\sigma_i)} |\sigma| \geq \max_{\sigma \in S_n^{(i-1)}} |\sigma|$, then $\min_{\sigma \in S'_n(\sigma_i)} |\sigma| \geq \max_{\sigma \in S_n^{(i-1)}(\sigma_i)} |\sigma|$ and $\min_{\sigma \in S'_n(\sigma_i)} |\sigma| \geq |\sigma_i| + 1$. Therefore under the previous condition the code $(S'_n \setminus S'_n(\sigma_i)) \cup S_n^{(i-1)}(\sigma_i)$ would be a strictly better symmetric fix-free code than S'_n for any choice of source probabilities, contradicting the assumption that $S'_n \in \mathbb{O}_n$. Hence,

$$\min_{\sigma \in \mathcal{N}_n(\sigma_i)} |\sigma| < \max_{\sigma \in S_n^{(i-1)}} |\sigma|. \quad (43)$$

$S_n^{(i)}$ consists of the smallest n elements of $S_n^{(i-1)} \cup \mathcal{N}_n(\sigma_i) \setminus \{\sigma_i\}$, so (43) implies that $\tilde{l}_n^{(i-1)} = \max_{\sigma \in S_n^{(i-1)}} |\sigma| \geq \max_{\sigma \in S_n^{(i)}} |\sigma| = \tilde{l}_n^{(i)}$. Hence, $l'_n = \tilde{l}_n^{(h)} \leq \tilde{l}_n^{(0)} = l_n$.

To begin our argument for the remainder of Theorem 14, recall our assumption that

$$S_n = S_n^{(0)} \xRightarrow{\sigma_1} S_n^{(1)} \xRightarrow{\sigma_2} S_n^{(2)} \xRightarrow{\sigma_3} \dots \xRightarrow{\sigma_k} S_n^{(h)} = S'_n$$

is a shortest sequence of codes in \mathbb{S}_n transforming S_n to S'_n via uses of the \Rightarrow operation.

Suppose $\{\pi_{1,1}, \pi_{2,1}, \dots, \pi_{k,1}\} = \{\sigma_1, \sigma_2, \dots, \sigma_h\} \cap S_n$ and the elements of $\{\sigma_1, \sigma_2, \dots, \sigma_h\} \setminus S_n$ each have a proper prefix in the set $\{\pi_{1,1}, \pi_{2,1}, \dots, \pi_{k,1}\}$. Then each string σ_ι , $\iota \in \{1, \dots, h\}$, can alternatively be labeled $\pi_{g,j}$, where

- if $\sigma_\iota \in S_n$, then $j = 1$ and $g = |\{\sigma_1, \sigma_2, \dots, \sigma_\iota\} \cap S_n|$, and
- if $\sigma_\iota \notin S_n$, then j is one more than the number of strings among $\{\sigma_1, \sigma_2, \dots, \sigma_\iota\}$ that have $\pi_{g,1}$ as a proper prefix.

Let γ_g be the number of strings, including $\pi_{g,1}$, among $\{\sigma_1, \sigma_2, \dots, \sigma_h\}$ which have $\pi_{g,1}$ as a prefix.

Let ρ_i , $i \in \{1, \dots, k\}$, be an arbitrary permutation of $\{1, \dots, k\}$. Then Theorem 8 implies that if $S'_n \in \mathbb{O}_n$, we can study the transformation from S_n to S'_n through any ordering of $\{\sigma_1, \sigma_2, \dots, \sigma_h\}$ of the form

$$\{\pi_{\rho_1,1}, \dots, \pi_{\rho_1,\gamma_{\rho_1}}, \pi_{\rho_2,1}, \dots, \pi_{\rho_2,\gamma_{\rho_2}}, \dots, \pi_{\rho_k,1}, \dots, \pi_{\rho_k,\gamma_{\rho_k}}\}. \quad (44)$$

We will use induction on k to show that the condition $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$ implies that $S'_n \notin \mathbb{O}_n$. For the basis step, Proposition 13 treats the case $k = 1$. For the inductive step, we assume the result is true when $k \leq \kappa$ and show that it is consequently true at $k = \kappa + 1$.

We will consider the possible transformations from S_n to S'_n using a permutation of $\{\sigma_1, \sigma_2, \dots, \sigma_h\}$ of the form (44). If $S'_n \in \mathbb{O}_n$, then by Theorem 8 we can define for $i \in \{1, \dots, \kappa + 1\}$ a sequence of symmetric fix-free codes $C_n^{(i,1)}, C_n^{(i,2)}, \dots, C_n^{(i,\gamma_i)} = I_n^{(i)} \in \mathbb{S}_n$ for which

$$S_n \xRightarrow{\pi_{i,1}} C_n^{(i,1)} \xRightarrow{\pi_{i,2}} \dots \xRightarrow{\pi_{i,\gamma_i}} C_n^{(i,\gamma_i)} = I_n^{(i)}.$$

For $1 \leq i \leq \kappa + 1$, let $(l_1^{(i)}, \dots, l_n^{(i)})$ denote the sorted and non-decreasing sequence of codeword lengths of $I_n^{(i)}$. If for any i , $\sum_{j=1}^n l_j \geq \sum_{j=1}^n l_j^{(i)}$, then the condition $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$ implies that $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j^{(i)}$. By the inductive hypothesis it follows from the transformation from $I_n^{(i)}$ to S'_n that $S'_n \notin \mathbb{O}_n$.

Therefore, assume for all $i \leq \kappa + 1$ that $\sum_{j=1}^n l_j^{(i)} > \sum_{j=1}^n l_j$. Define λ_i as the smallest integer for which

$$I_n^{(i)} \subseteq S_n \cup \mathcal{N}_{\lambda_i}^* (\pi_{i,1}) \setminus \{\{\pi_{i,1}\} \cup \tilde{S}_n^{\lambda_i}\}.$$

Let

$$\begin{aligned} D_i &= S_n \setminus I_n^{(i)} \\ D'_i &= I_n^{(i)} \setminus S_n \\ m_i &= |D_i| = |D'_i|. \end{aligned}$$

Let $(d_{i,1}, \dots, d_{i,m_i})$ and $(d'_{i,1}, \dots, d'_{i,m_i})$ be the sorted and non-decreasing sequences of codeword lengths of D_i and D'_i , respectively. Then by (41) and (42) we have

$$\sum_{j=1}^t d'_{i,j} \geq \sum_{j=1}^t d_{i,j}, \quad 1 \leq t \leq m_i, \quad (45)$$

and we also have

$$|\pi_{i,1}| = d_{i,1} < \lambda_i + 1 \leq d_{i,2} \leq \dots \leq d_{i,m_i} \quad (46)$$

$$d_{i,1} + 1 \leq d'_{i,1} \leq d'_{i,2} \leq \dots \leq d'_{i,m_i} \leq \lambda_i \quad (47)$$

Define μ as the smallest integer for which

$$S'_n \subseteq S_n \cup \left[\bigcup_{i=1}^{\kappa+1} \mathcal{N}_\mu^*(\pi_{i,1}) \right] \setminus \left[\bigcup_{i=1}^{\kappa+1} \{\pi_{i,1}\} \cup \tilde{S}_n^\mu \right].$$

Observe that

$$\mu \leq \min \{\lambda_1, \dots, \lambda_{\kappa+1}\}. \quad (48)$$

Let $m = |S_n \setminus S'_n|$, and let $(\delta_1, \dots, \delta_m)$ and $(\delta'_1, \dots, \delta'_m)$ be the sorted and non-decreasing sequences of codeword lengths of $S_n \setminus S'_n$ and $S'_n \setminus S_n$, respectively. If S'_n is a candidate for membership in \mathbb{O}_n and we are studying part of a shortest transformation from R_n to S'_n , then because $\delta_1, \dots, \delta_{\kappa+1}$ are the ordered lengths of $\pi_{1,1}, \dots, \pi_{\kappa+1,1}$, it follows that

$$\delta_1 \leq \delta_2 \leq \dots \leq \delta_{\kappa+1} < \mu + 1 \leq \delta_{\kappa+2} \leq \dots \leq \delta_m \quad (49)$$

$$\delta_1 + 1 \leq \delta'_1 \leq \delta'_2 \leq \dots \leq \delta'_m \leq \mu \quad (50)$$

As in the proof of Proposition 13, we can argue that $S'_n \notin \mathbb{O}_n$ if $\sum_{i=1}^t \delta'_i \geq \sum_{i=1}^t \delta_i$ for all $t \leq m$. The condition $\sum_{i=1}^n l'_i \geq \sum_{i=1}^n l_i$ here implies that $\sum_{i=1}^m \delta'_i \geq \sum_{i=1}^m \delta_i$. To establish the remaining $m - 1$ inequalities we consider three cases:

- 1) $t = 1$: We know that $\delta'_1 \geq \delta_1 + 1$.
- 2) $2 \leq t \leq \kappa + 1$: Starting from $t = 1$ we will sequentially map each $t \leq \kappa + 1$ into a different ordered pair $(i(t), j(t))$ satisfying $\delta'_t = d'_{i(t), j(t)}$ as follows. If there are multiple unchosen pairs $(i(t), j(t))$ which satisfy the equality then we select the one with minimum $j(t)$ and then, if necessary, minimum $i(t)$. Let

$$\begin{aligned} \mathcal{I}_t &= \{i : \tau \rightarrow (i, j) \text{ for some } \tau \leq t\} \\ j_t(i) &= |\{j : \tau \rightarrow (i, j) \text{ for some } \tau \leq t\}| \end{aligned}$$

Then

$$\begin{aligned} \sum_{a=1}^t \delta'_a &= \sum_{i \in \mathcal{I}_t} \sum_{j=1}^{j_t(i)} d'_{i,j} \stackrel{(a)}{\geq} \sum_{i \in \mathcal{I}_t} \sum_{j=1}^{j_t(i)} d_{i,j} \stackrel{(b)}{\geq} \sum_{i \in \mathcal{I}_t} \left(d_{i,1} + \sum_{j=2}^{j_t(i)} (\lambda_i + 1) \right) \\ &\stackrel{(c)}{\geq} \sum_{i \in \mathcal{I}_t} \left(d_{i,1} + \sum_{j=2}^{j_t(i)} \mu \right) \\ &\stackrel{(d)}{\geq} \sum_{a=1}^t \delta_a. \end{aligned}$$

Here (a) follows from (45), (b) follows from (46), (c) follows from (48) and (d) follows from (49) and the assumption that $t \leq \kappa + 1$.

- 3) $\kappa+2 \leq t \leq m-1$: We are given $\sum_{i=1}^m \delta'_i \geq \sum_{i=1}^m \delta_i$ or, equivalently, $\sum_{i=1}^{\kappa+1} (\delta'_i - \delta_i) \geq \sum_{i=\kappa+2}^m (\delta_i - \delta'_i)$. (49) and (50) imply that for $i \geq \kappa+2$,

$$\delta_i \geq \delta'_i + 1.$$

Hence for $t \geq \kappa+2$,

$$\sum_{i=1}^t (\delta'_i - \delta_i) \geq \sum_{i=t+1}^m (\delta_i - \delta'_i) \geq 0.$$

Thus the condition $\sum_{i=1}^n l'_i \geq \sum_{i=1}^n l_i$ here implies that $S'_n \notin \mathbb{O}_n$. \square

Theorem 14 shows conditions for which the n inequalities of Lemma 12 can be reduced to one. We next show that if by an application of Proposition 13 or Theorem 14 we determine that $S'_n \notin \mathbb{O}_n$, then we can automatically conclude that certain related codes also are not members of \mathbb{O}_n . We have the following result.

Theorem 15: Suppose that the codes $S_n, S'_n, C_n \in \mathbb{S}_n$, that S_n is in a shortest transformation from R_n to S'_n through a sequence of \Rightarrow operations, and that S'_n is in a shortest transformation from R_n to C_n through a sequence of \Rightarrow operations. Let (l_1, l_2, \dots, l_n) and $(l'_1, l'_2, \dots, l'_n)$ be the sorted and non-decreasing sequences of codeword lengths of S_n and S'_n , respectively. Suppose that $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$. If the portion of the shortest transformation from S_n to S'_n satisfies either

- $S_n \xRightarrow{\pi_1} S'_n$ or
- there is a sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(h)} \in \mathbb{S}_n$ for some $h \geq 2$ with

$$S_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_h} S_n^{(h)} = S'_n$$

and with π_1 being a prefix of π_i for $i \geq 2$,

and the portion of the shortest transformation from S'_n to C_n can be described for some $\eta \geq 1$ by

$$S'_n \xRightarrow{\sigma_1} C_n^{(1)} \xRightarrow{\sigma_2} C_n^{(2)} \xRightarrow{\sigma_3} \dots \xRightarrow{\sigma_\eta} C_n^{(\eta)} = C_n$$

with π_1 not being a prefix of σ_i for $1 \leq i \leq \eta$, then $C_n \notin \mathbb{O}_n$.

Proof: Following the notation introduced in the proof of Proposition 13, let

$$\begin{aligned} D &= S_n \setminus S'_n = \{\tilde{s}_1, \dots, \tilde{s}_m\}, \\ D' &= S'_n \setminus S_n = \{s'_1, \dots, s'_m\}, \\ m &= |D| = |D'|, \end{aligned}$$

and let (d_1, \dots, d_m) and (d'_1, \dots, d'_m) respectively denote the sorted and non-decreasing sequences of codeword lengths of D and D' . To arrive at a contradiction, suppose $C_n \in \mathbb{O}_n$. Then by Lemma 7, π_1 must be a prefix of some element of C_n , and therefore

$$D' \cap C_n \neq \emptyset.$$

Suppose $|D' \cap C_n| = k$. By the definition of the \Rightarrow operation, $D' \cap C_n = \{s'_1, \dots, s'_k\}$.

From the proof of Proposition 13 we saw that the condition $\sum_{i=1}^n l'_i \geq \sum_{i=1}^n l_i$ implies that $\sum_{j=1}^t d'_j \geq \sum_{j=1}^t d_j$ for all $1 \leq t \leq m$. Therefore, the sequence of sorted and non-decreasing lengths of the strings in $C_n \cup \{\tilde{s}_1, \dots, \tilde{s}_k\} \setminus \{s'_1, \dots, s'_k\}$ dominates the sequence of sorted and non-decreasing codeword lengths of C_n . To complete the

proof it suffices to show that $C_n \cup \{\tilde{s}_1, \dots, \tilde{s}_k\} \setminus \{s'_1, \dots, s'_k\} \in \mathbb{S}_n$. By the definition of the \Rightarrow operation, we have that $D \cap C_n = \emptyset$. Furthermore, for $1 \leq i \leq \eta$, $\sigma_i \notin D$ because either $\sigma_i \in S'_n$ or $\sigma_i \in \mathcal{N}_n^*(\sigma_j)$ for some $j < i$ with $\sigma_j \in S'_n$. Hence $C_n \cup \{\tilde{s}_1, \dots, \tilde{s}_k\} \setminus \{s'_1, \dots, s'_k\} \in \mathbb{S}_n$. \square

Recall that $R_n = \{s_1, s_2, \dots, s_n\}$. We have the following result.

Corollary 16: Let C^{prefix} be the set of palindromes (not including 1) which are proper prefixes of at least one codeword in $C_n \in \mathbb{O}_n$. For $i \geq n/2$, $s_i \notin C^{\text{prefix}}$.

Proof: For $i \geq (n+2)/2$, $\min_{\sigma \in \mathcal{N}(s_i)} |\sigma| = 2i - 1 \geq n + 1$, so the $\xRightarrow{s_i}$ operation would not produce a code in \mathbb{S}_n . If n is odd, then the shortest two palindromes which have $s_{[(n+1)/2]}$ as a proper prefix have lengths n and $n+1$. If n is even, then the shortest two palindromes which have $s_{[n/2]}$ as a proper prefix have lengths $n-1$ and n . In either of these cases it is better to keep $s_{[n/2]}$ or $s_{[(n+1)/2]}$ as a codeword than to turn it into a proper prefix of one. \square

Observe that for a string σ and its bitwise complement $\bar{\sigma}$, the lengths of strings in $\mathcal{N}_n(\sigma)$ will match those of their bitwise complements in $\mathcal{N}_n(\bar{\sigma})$. Therefore, the previous result implies that if $0 \in C^{\text{prefix}}$, then for $i \geq n/2$, $\bar{s}_i \notin C^{\text{prefix}}$. More generally if a code S_n contains σ and $\bar{\sigma}$, then one can impose an ordering on them for C^{prefix} and thereby reduce the number of strings to be considered for replacement at the next step. Furthermore, we immediately obtain the following extension to Theorem 15.

Corollary 17: Suppose that the codes $S_n, S'_n \in \mathbb{S}_n$ and that S_n is in a shortest transformation from R_n to S'_n through a sequence of \Rightarrow operations. Let (l_1, l_2, \dots, l_n) and $(l'_1, l'_2, \dots, l'_n)$ be the sorted and non-decreasing sequences of codeword lengths of S_n and S'_n , respectively. Suppose that $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$. If the portion of the shortest transformation from S_n to S'_n satisfies either

- $S_n \xRightarrow{\pi_1} S'_n$ or
- there is a sequence of symmetric fix-free codes $S_n^{(1)}, S_n^{(2)}, \dots, S_n^{(h)} \in \mathbb{S}_n$ for some $h \geq 2$ with

$$S_n = S_n^{(0)} \xRightarrow{\pi_1} S_n^{(1)} \xRightarrow{\pi_2} S_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_h} S_n^{(h)} = S'_n$$

and with π_1 being a prefix of π_i for $i \geq 2$,

and if $\bar{\pi}_1 \in S_n$, then the code S''_n defined by

$$S_n \xRightarrow{\bar{\pi}_1} \hat{S}_n^{(1)} \xRightarrow{\pi_2} \hat{S}_n^{(2)} \xRightarrow{\pi_3} \dots \xRightarrow{\pi_h} \hat{S}_n^{(h)} = S''_n$$

is not an element of \mathbb{O}_n . Furthermore, for $\eta \geq 1$ any code \hat{C}_n related to S''_n by a transformation of the form

$$S''_n \xRightarrow{\sigma_1} \hat{C}_n^{(1)} \xRightarrow{\sigma_2} \hat{C}_n^{(2)} \xRightarrow{\sigma_3} \dots \xRightarrow{\sigma_\eta} \hat{C}_n^{(\eta)} = \hat{C}_n$$

for which $\bar{\pi}_1$ not being a prefix of σ_i for $1 \leq i \leq \eta$ satisfies $\hat{C}_n \notin \mathbb{O}_n$.

There would be a further simplification in using these ideas to generate all optimal symmetric fix-free codes if the following conjecture holds:

Conjecture 18: Suppose that the codes $S_n, S'_n, C_n \in \mathbb{S}_n$, that S_n is in a shortest transformation from R_n to S'_n through a sequence of \Rightarrow operations, and that S'_n is in a shortest transformation from R_n to C_n through a sequence of \Rightarrow operations. Let (l_1, l_2, \dots, l_n) and $(l'_1, l'_2, \dots, l'_n)$ be the sorted and non-decreasing sequences of codeword lengths of S_n and S'_n , respectively. Suppose that $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$. If for

some $\eta \geq 1$ a shortest transformation from S_n to C_n can be described $S_n \xrightarrow{\pi} S'_n \xrightarrow{\sigma_1} C_n^{(1)} \xrightarrow{\sigma_2} C_n^{(2)} \xrightarrow{\sigma_3} \dots \xrightarrow{\sigma_\eta} C_n^{(\eta)} = C_n$, then $C_n \notin \mathbb{O}_n$. If in addition $\bar{\pi} \in S_n$ and $S_n \xrightarrow{\bar{\pi}} S''_n$, then $S''_n \notin \mathbb{O}_n$ and S''_n is not in any shortest transformation from R_n to a code in \mathbb{O}_n .

If this conjecture is true, then at each code S_n generated as a candidate member of \mathbb{O}_n we need only consider additional transformations involving codewords which when replaced will result in codes with smaller sums of codeword lengths than that of S_n . Furthermore we obtain constraints on C^{prefix} which may result in other reductions to our search space for optimal codes. However, while this conjecture is open, one way to effectively use Theorems 14 and 15 is to establish for each code S_n and string π whether or not the conditions $S_n \xrightarrow{\pi} S'_n$ and $\sum_{j=1}^n l'_j \geq \sum_{j=1}^n l_j$ imply that (1) S_n also has a sum of codeword lengths which is at most that of any code $C_n \in \mathbb{S}_n$ given by $S'_n \xrightarrow{\sigma_1} C_n^{(1)} \xrightarrow{\sigma_2} C_n^{(2)} \xrightarrow{\sigma_3} \dots \xrightarrow{\sigma_\eta} C_n^{(\eta)} = C_n$, where π is a prefix of σ_i for each $1 \leq i \leq \eta$ or (2) the preceding sequence of code transformations is not associated with a non-increasing sequence of maximum codeword lengths. If these latter constraints can be verified for a given code S_n and string π , then it can be concluded that S'_n is not in any shortest transformation from R_n to any code in \mathbb{O}_n ; as we indicated earlier, this places restrictions on C^{prefix} for optimal codes.

We have mentioned earlier that $R_n \in \mathbb{O}_n$ for $n \geq 3$. This is the only optimal symmetric fix-free code for $n = 3$ and $n = 4$. We next describe some of the other codes in \mathbb{D}_n for $n \geq 5$.

Theorem 19: Let (l_1, l_2, \dots, l_n) be the sorted and non-decreasing sequence of codeword lengths for a code $S_n \in \mathbb{S}_n$ satisfying $R_n \Rightarrow S_n$. Then $S_n \in \mathbb{D}_n$ if $\sum_{i=1}^n l_i < n(n+1)/2$.

Proof: Assume j is the index for which $S_n \subseteq (R_n \setminus \{s_j\}) \cup \mathcal{N}_n(s_j)$. To arrive at a contradiction, suppose that there is a code $C_n = \{c_1, \dots, c_n\} \in \mathbb{S}_n$ which differs from both S_n and its complementary code, satisfies $|c_1| \leq |c_2| \leq \dots \leq |c_n|$, and has the property that

$$\sum_{i=1}^{\kappa} |c_i| \leq \sum_{i=1}^{\kappa} l_i, \text{ for all } \kappa \in \{1, \dots, n\}. \quad (51)$$

Since $\sum_{i=1}^n |c_i| < n(n+1)/2$, it follows that $C_n \neq R_n$. By Lemma 3, each codeword of C_n has a prefix in $R_n = \{s_1, \dots, s_n\}$. Since $C_n \neq R_n$, there exists ι and γ such that s_γ is a proper prefix of c_ι . Let k be the smallest index for which $s_k \notin C_n$. Therefore, since the shortest string of $\mathcal{N}_n(s_\gamma)$ has length $\max\{2\gamma - 1, 2\}$,

$$|c_\iota| \geq \max\{2\gamma - 1, 2\} \geq \max\{2k - 1, 2\}. \quad (52)$$

Since $C_n \in \mathbb{S}_n$ it follows that $2k - 1 \leq n$.

We next show that the first $\max\{2k - 2, 1\}$ sorted and non-decreasing codeword lengths of C_n satisfy

$$|c_i| = i, \quad i \leq k - 1 \quad (53)$$

$$|c_i| \geq i + 1, \quad k \leq i \leq \max\{2k - 2, 1\}. \quad (54)$$

If $k = 1$, then $0, 1 \notin C_n$, so $|c_1| \geq 2$. If $k \geq 2$, then (53) holds because $\{s_1, \dots, s_{k-1}\} \subset C_n$. By the Kraft inequality, $|c_k| \geq k - 1$ with strict inequality since $(1, 2, \dots, k - 1, k - 1)$ is not a feasible sequence of codeword lengths among symmetric fix-free codes.

If $|c_k| = k$, then $\{s_1, \dots, s_{k-1}\} \subset C_n$ implies $s_k \in C_n$, which contradicts the definition of k . Therefore $|c_k| \geq k + 1$. For $k \geq 3$, $k + 1 \leq i \leq 2k - 2$, suppose that (54) is not always true. Then there is a smallest index $t \in \{k + 1, \dots, 2k - 2\}$ such that $|c_t| \leq t$. Since $|c_{t-1}| \geq t$, we have

$$|c_{t-1}| = |c_t| = t \leq 2k - 2. \quad (55)$$

We next show that $c_t \in R_n$. To arrive at a contradiction, suppose that $s_q \in R_n$ is a proper prefix of c_t . By the same argument as for (52), we have that $|c_t| \geq 2k - 1$, which contradicts (55). Hence, $c_t \in R_n$. The same argument implies that $c_{t-1} \in R_n$, but $|c_t| \neq |c_{t-1}|$ for different elements of R_n . Thus, (54) follows because (55) is false.

There are three cases to consider to establish the result:

- 1) $1 \leq k < j \leq n$: Since R_k is a subset of S_n , it follows that $l_i \leq i$ for $i \leq k$. Therefore, by (53) and (54), $\sum_{i=1}^k l_i < \sum_{i=1}^k |c_i|$, which contradicts (51).
- 2) $1 \leq k = j \leq n$: We have $1, s_j \notin C_n$, so each codeword $c_i \in C_n$ has a prefix $w_i \in (R_n \setminus \{s_j\}) \cup \mathcal{N}_n(s_j)$. Furthermore, S_n consists of the shortest n strings in $(R_n \setminus \{s_j\}) \cup \mathcal{N}_n(s_j)$. Therefore, for $i = 1$, $|c_1| \geq |w_1| \geq l_1$, and so (51) implies that $c_1 = w_1$. Next suppose that there is an index $\lambda \in \{1, \dots, n - 1\}$ such that $c_i = w_i$ for all $i \leq \lambda$. Observe that if $w_{\lambda+1} \in \{w_1, \dots, w_\lambda\} = \{c_1, \dots, c_\lambda\}$, then $\{c_1, \dots, c_\lambda, c_{\lambda+1}\}$, does not satisfy the prefix condition and cannot be a symmetric fix-free code. Hence, $w_{\lambda+1} \notin \{w_1, \dots, w_\lambda\}$. Therefore, $\{w_1, \dots, w_\lambda, w_{\lambda+1}\}$ is a subset of $(R_n \setminus \{s_j\}) \cup \mathcal{N}_n(s_j)$ with $\lambda + 1$ distinct elements. Thus, $\sum_{i=1}^{\lambda+1} |c_i| \geq \sum_{i=1}^{\lambda+1} |w_i| \geq \sum_{i=1}^{\lambda+1} l_i$. It follows from (51) that $\sum_{i=1}^{\lambda+1} |c_i| = \sum_{i=1}^{\lambda+1} l_i$. By induction $(|c_1|, |c_2|, \dots, |c_n|) = (l_1, \dots, l_n)$, which contradicts our earlier assumption.
- 3) $1 \leq j < k \leq n$: Let v be the shortest element of $\mathcal{N}_n(s_j)$. Define the code $B_{2k-1} = (R_{2k-1} \setminus \{s_j\}) \cup \{v\} \subseteq (R_n \setminus \{s_j\}) \cup \mathcal{N}_n(s_j)$. Let b_{sum} be the sum of lengths of the elements of B_{2k-1} . Therefore,

$$\sum_{i=1}^{2k-1} l_i = b_{sum} \quad (56)$$

$$\begin{aligned} &= 2k^2 - k - j + \max\{2j - 1, 2\} \\ &\leq 2k^2 - 1 \end{aligned} \quad (57)$$

$$\begin{aligned} &= \sum_{i=1}^{k-1} i + \left(\sum_{i=k}^{2k-2} (i + 1) \right) + (2k - 1) \\ &\leq \sum_{i=1}^{2k-1} |c_i|, \end{aligned} \quad (58)$$

by (53), (54), and the fact that $|c_{2k-1}| \geq |c_{2k-2}| \geq 2k - 1$. The only way for (58) to be consistent with (51) is for (56), (57), and (58) all to be equalities. In order for (57) to be an equality, $j = 1$ and $k = 2$. Since $j = 1$, $S_n = \{00, 11, 010, 101, \dots\}$. For (58) to be an equality, $|c_1| = 1$, $|c_2| = 3$, $|c_3| = 3$. Recall that we assume that $n \geq 5$. Since $c_1 = 0$, $c_2, c_3 \in \{101, 111\}$, it follows that $|c_4| \geq 4$. However, for these choices of S_n and C_n we find that $\sum_{i=1}^4 l_i = 10 < 11 \leq \sum_{i=1}^4 |c_i|$, which again contradicts (51).

Since each way of constructing the symmetric fix-free code C_n results in a violation of an assumption, we find that $S_n \in \mathbb{D}_n$. \square

One can use the experimental results of [13] to show that R_n and the optimal codes of Theorem 19 make up all of the optimal codes for $n \leq 10$.

Our last technical result establishes a special case of Conjecture 18.

Theorem 20: Suppose symmetric fix-free codes S'_n and C_n are related to each other and to R_n by $R_n \xrightarrow{s_i} S'_n \xrightarrow{\sigma} C_n$, and suppose $S'_n \notin \mathbb{D}_n$. Then $C_n \notin \mathbb{O}_n$.

Proof: The case where $\sigma = s_\gamma$ for $\gamma \neq \iota$ is established by Theorem 15. Therefore we will assume that $\sigma \in \mathcal{N}_n(s_\iota)$. As usual, let l'_n denote the maximum codeword length of S'_n . We will prove the result by arguing that $\min_{\psi \in \mathcal{N}(\sigma)} |\psi| > l'_n$. Because of the structure of s_ι , it is simple to establish that $|\sigma| \geq 2\iota - 1$ and $\min_{\psi \in \mathcal{N}(\sigma)} |\psi| \geq 3\iota - 2$. Therefore it suffices to show

$$3\iota - 2 > l'_n. \quad (59)$$

As in earlier proofs, let

$$m = |S'_n \setminus R_n| = |R_n \setminus S'_n|. \quad (60)$$

Let 0^r denote a string of r zeroes and let θ^ρ denote a palindrome of length ρ . The shortest elements of $\mathcal{N}_n(s_\iota)$ are of the form $10^{\iota-2}10^{\iota-2}1$, $s_\iota s_\iota$, $s_\iota \theta^1 s_\iota$, $s_\iota \theta^2 s_\iota$, \dots , $s_\iota \theta^{\iota-3} s_\iota$. For $0 \leq \rho \leq \iota - 3$, every palindrome θ^ρ satisfies $s_\iota \theta^\rho s_\iota \in \mathcal{N}(s_\iota)$. Since there are $2^{\lfloor (\rho+1)/2 \rfloor}$ palindromes of length ρ , if $m \leq \sum_{t=0}^{\iota-2} 2^{\lfloor t/2 \rfloor}$, then we have a complete description of $S'_n \setminus R_n$.

In Corollary 16 we showed the desired result when $\iota \geq n/2$. Therefore, we need only consider the case where $\iota \leq (n-1)/2$. Suppose that for some $k \geq 1$,

$$l'_n = 2\iota + k. \quad (61)$$

It follows from (59) and (61) that we would like to show

$$k \leq \iota - 3. \quad (62)$$

Note that the preceding condition would also imply that the longest codeword of $S'_n \setminus R_n$ is of the form $s_\iota \theta^k s_\iota$ for an arbitrary length- k palindrome θ^k and that we could completely describe $S'_n \setminus R_n$.

By the definition of the \Rightarrow operation, $R_n \setminus S'_n = \{s_\iota, s_{2\iota+k+1}, s_{2\iota+k+2}, \dots, s_n\}$, and the sum of the lengths of these words is

$$\iota + \sum_{i=2\iota+k+1}^n i = \iota + n(m-1) - \frac{(m-1)(m-2)}{2}. \quad (63)$$

In order to find k , we wish to have

$$|\mathcal{N}_{2\iota+k-1}(s_\iota)| < m \leq |\mathcal{N}_{2\iota+k}(s_\iota)|. \quad (64)$$

If (62) holds and k is odd, then k satisfies

$$\sum_{t=0}^k 2^{\lfloor t/2 \rfloor} = 2^{(k+3)/2} - 2 < m \leq 2^{(k+3)/2} + 2^{(k+1)/2} - 2 = \sum_{t=0}^{k+1} 2^{\lfloor t/2 \rfloor}. \quad (65)$$

If (62) holds and k is even, then k satisfies

$$\sum_{t=0}^k 2^{\lfloor t/2 \rfloor} = 2^{(k+2)/2} + 2^{k/2} - 2 < m \leq 2^{(k+4)/2} - 2 = \sum_{t=0}^{k+1} 2^{\lfloor t/2 \rfloor}. \quad (66)$$

Observe that if (62) holds, then the sum of the codeword lengths over the set $S'_n \setminus R_n$ is

$$m(2\iota-1) + \sum_{t=0}^k t \cdot 2^{\lfloor t/2 \rfloor} + (k+1) \left(m - \sum_{t=0}^k 2^{\lfloor t/2 \rfloor} \right) = m(2\iota+k) - \sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor}. \quad (67)$$

Since $S'_n \notin \mathbb{D}_n$, we have that the sum of codeword lengths over $R_n \setminus S'_n$ is at most the sum of codeword lengths over $S'_n \setminus R_n$. Hence (67) and (63) imply that

$$\iota + n(m-1) - \frac{(m-1)(m-2)}{2} \leq m(2\iota+k) - \sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor}. \quad (68)$$

Since

$$m = |R_n \setminus S'_n| = |\{s_\iota, s_{2\iota+k+1}, s_{2\iota+k+2}, \dots, s_n\}| = n+1-2\iota-k, \quad (69)$$

the condition (68) can be rewritten

$$\frac{n}{2} \geq \frac{m^2 - k - 1}{2} + \sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor}. \quad (70)$$

Because of (69), the condition (62) that we wish to establish is equivalent to

$$\frac{n}{2} \geq \frac{m+3k+5}{2}. \quad (71)$$

Therefore, to demonstrate (71) it sufficient to show that

$$\frac{m^2 - k - 1}{2} + \sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor} \geq \frac{m+3k+5}{2}$$

or

$$\frac{m^2 - m}{2} - 2k - 3 + \sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor} \geq 0. \quad (72)$$

If k is odd, then $\sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor} = 7 \cdot 2^{(k+1)/2} - 2k - 9$, and we wish to verify if

$$\frac{m^2 - m}{2} + 7 \cdot 2^{(k+1)/2} - 4k - 12 \geq 0 \quad (73)$$

when m satisfies (65). The expression $m^2 - m$ is minimized when $m = 2^{(k+3)/2} - 1$, and for this m the left-hand side of (73) is $2^{k+2} + 4 \cdot 2^{(k+1)/2} - 4k - 11 \geq 1$ for $k \geq 1$. If $k \geq 2$ is even, one can show that $\sum_{t=0}^k (k+1-t) \cdot 2^{\lfloor t/2 \rfloor} = 10 \cdot 2^{k/2} - 2k - 9$, and we wish to assess if

$$\frac{m^2 - m}{2} + 10 \cdot 2^{k/2} - 4k - 12 \geq 0 \quad (74)$$

when m satisfies (66). The expression $m^2 - m$ is minimized when $m = 3 \cdot 2^{k/2} - 1$, and for this m the left-hand side of (74) is $9 \cdot 2^{k-1} + 5.5 \cdot 2^{k/2} - 4k - 11 \geq 10$ for $k \geq 2$.

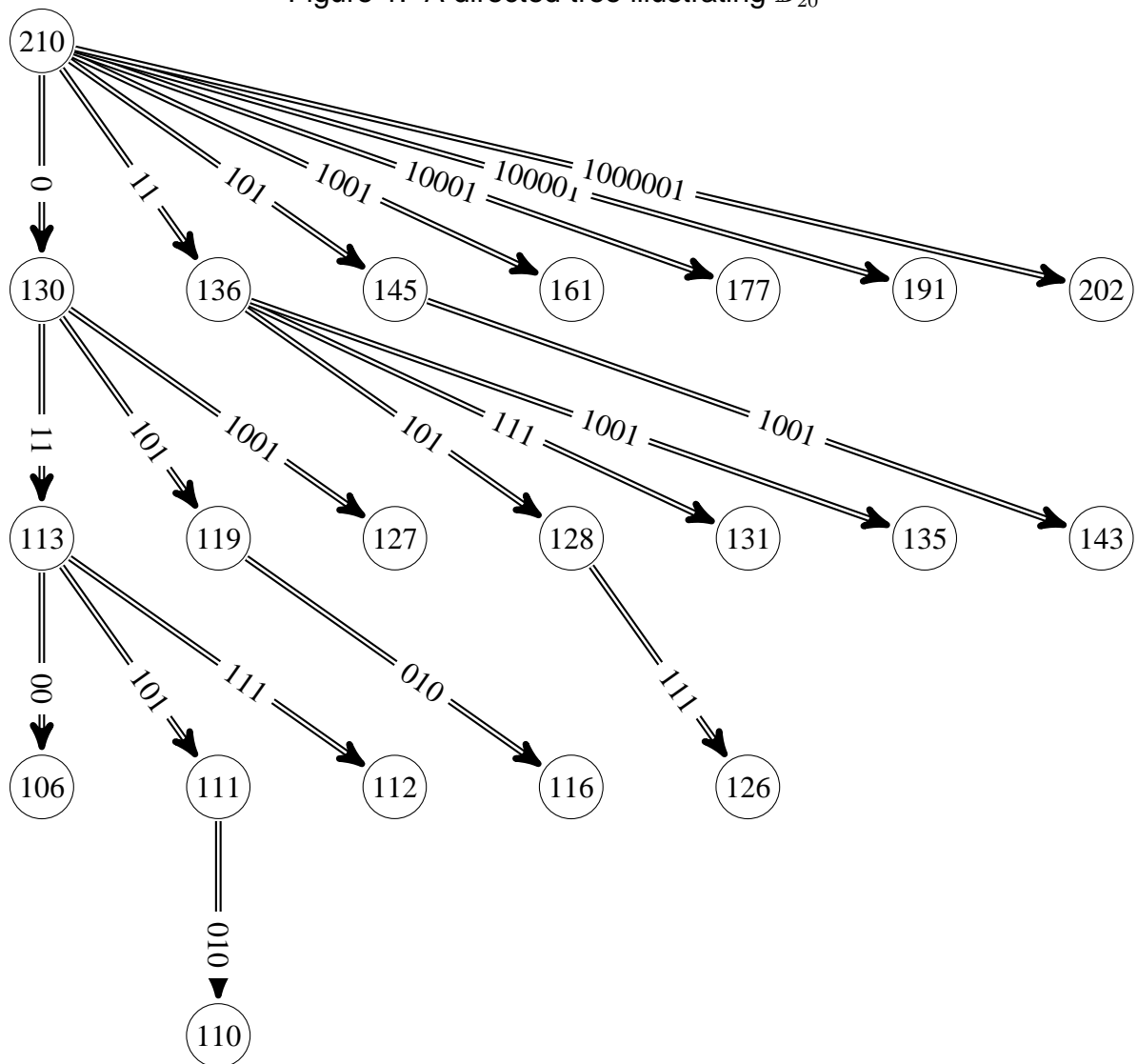
Since (72) holds for all $k \geq 1$, the result follows. \square

In Figure 1, we illustrate the tree of all 21 codes in \mathbb{D}_{20} . The numbers within the vertices represent the sum of codeword lengths for the corresponding code. The strings

labeling the edges represent the codeword removed to go from a code to the next one. The codelength sequences discussed in [11] form a lattice instead of a tree. Furthermore in [11] the codelength sequence with minimum sum was the furthest away from that corresponding to the most imbalanced code, while this is not the case here. However, both here and in [11] the most imbalanced (optimal) code of the class being studied had a central role in a mathematical analysis of optimal codes.

Acknowledgment. The authors were supported by NSF Grant No. CCF-1017303.

Figure 1. A directed tree illustrating \mathbb{D}_{20}



References

- [1] N. Abedini, S. P. Khatri, and S. A. Savari, "A SAT-based scheme to determine optimal fix-free codes," *Proc. of the 2010 IEEE Data Compression Conference*, pp. 169-178, Snowbird, Utah, March 2010.
- [2] R. Bauer and J. Hagenauer, "Iterative source/channel-decoding using reversible variable length codes," *Proc. IEEE Data Compression Conference*, pp. 93-102, Snowbird Utah, March 2000.

- [3] R. Bauer and J. Hagenauer, "On variable length codes for iterative source/channel-decoding," *Proc. IEEE Data Compression Conference*, pp. 273–282, Snowbird Utah, March 2001.
- [4] J. Berstel and D. Perrin, *Theory of Codes*, Orlando: Academic Press, 1985.
- [5] A. S. Fraenkel and S. T. Klein, "Bidirectional Huffman coding," *Comp. J.*, vol. 33, pp. 296-307, 1990.
- [6] E. N. Gilbert, "Codes based on inaccurate source probabilities," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 304-314, 1971.
- [7] E. N. Gilbert and E. F. Moore, "Variable-length binary encodings," *Bell Sys. Tech. J.*, vol. 38, pp. 933-967, July 1959.
- [8] Y. M. Huang, T. Y. Wu, and Y. S. Han, "An A^* -based algorithm for constructing reversible variable length codes with minimum average codeword length," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3175-3185, Nov. 2010.
- [9] K. Laković and J. Villasenor, "An algorithm for construction of efficient fix-free codes," *IEEE Commun. Letters*, vol. 7, pp. 391–393, Aug. 2003.
- [10] D. Marpe, T. Wiegand, and G. J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134-143, Aug. 2006.
- [11] D. S. Parker and P. Ram, "The construction of Huffman codes is a submodular ("convex") optimization problem over a lattice of binary trees," *SIAM J. Comput.*, vol. 28, no. 5, pp. 1875-1905, 1999.
- [12] S. A. Savari, "On optimal reversible-variable-length codes," *Proc. 2009 Information Theory and Applications Workshop*, pp. 311-317, La Jolla, CA, February 10, 2009.
- [13] S. A. Savari, S. M. H. Tabatabaei Yazdi, N. Abedini, and S. P. Khatri, "On optimal and achievable fix-free codes," *IEEE Trans. Inform. Theory*, vol. 58, no. 8, pp. 5112-5129, August 2012.
- [14] M. P. Schützenberger, "On an application of semi-group methods to some problems in coding," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 47–60, Sept. 1956.
- [15] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.
- [16] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Trans. Commun.*, vol. 43, pp. 158–162, Feb.-Apr. 1995.
- [17] C.-W. Tsai and J.-L. Wu, "On constructing the Huffman-code-based reversible variable-length codes," *IEEE Trans. Commun.*, vol. 49, pp. 1506–1509, Sept. 2001.
- [18] C.-W. Tsai and J.-L. Wu, "Modified symmetrical reversible variable-length codes and its theoretical bounds," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2543–2548, Sept. 2001.
- [19] H. W. Tseng and C. C. Chang, "Construction of symmetrical reversible variable length codes using backtracking," *Computer J.*, vol. 46, no. 1, pp. 100-105, 2003.
- [20] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

Appendix

Table 1 shows the exact number of different sorted and ascending codelength sequences for Huffman codes (i.e., binary prefix condition codes which satisfy the Kraft inequality with equality) and an upper bound for the counterpart for optimal symmetric fix-free codes with n words based on the number of dominant codelength sequences when $n \leq 30$. The numbers for the Huffman code are taken from [6], and the numbers for dominant length sequences for symmetric fix-free codes come from [1], [13].

Table 1. Number of (Sorted and Nondecreasing) Dominant Codelength Sequences over a Binary Code Alphabet

n	Huffman	Symmetric
2	1	1
3	1	1
4	2	1
5	3	2
6	5	2
7	9	3
8	16	3
9	28	4
10	50	4
11	89	6
12	159	6
13	285	8
14	510	11
15	914	11
16	1639	13
17	2938	13
18	5269	17
19	9451	18
20	16952	21
21	30410	22
22	54555	24
23	97871	26
24	175588	29
25	315016	32
26	565168	34
27	1013976	36
28	1819198	42
29	3263875	43
30	5855833	46